

Translators and Machine Translation. Book of Presentations.



**2nd International T3L Conference: Tradumatics, Translation Technologies & Localisation
“Translators and machine translation”**

**9th International Conference on Translation, Department of Translation, Interpreting and
East Asian Studies, Universitat Autònoma de Barcelona**



10-11 October 2016

Faculty of Translation and Interpreting, Universitat Autònoma de Barcelona



Organized by the Tradumàtica Research Group.



Acknowledgment

We would like to thank the presenters, who have given permission to include their presentations in this book of presentations.

Funding

This book has been funded by the project Proyectos de traducción con Traducción Automática Estadística y Posedición (ProjecTA) [Translation Projects with Statistical Machine Translation and Postediting]. Ref: FFI2013-46041-R. Funded by the Ministry of Economy and Competitiveness of Spain.

License

Creative Commons Attribution-NonCommercial 4.0 International Public License.



Recommended citation

Tradumàtica Research Group (2016). *Translators and Machine Translation: Book of presentations*. Bellaterra: Department of Translation, Interpreting and East Asian Studies. DOI: 10.13140/RG.2.2.12770.12484.
<http://ddd.uab.cat/pub/lilibres/2016/168388/tramactra_a2016.pdf>

TABLE OF CONTENTS

This document contains the abstracts and presentations for the conference.

Plenary sessions	5
New uses of machine translation in the translation workstation <i>Mikel Forcada</i>	6
Panorama 2020: which will be the most sought-after professional profiles in the translation marketplace? <i>Olga Blasco</i>	59
Conferences	98
The quality of post-edited machine translation: what's it to you? <i>Gys-Walt van Egdom</i>	99
Post-editing suboptimal machine translation: why, when and how? <i>Nora Aramberri</i>	108
A Description of Post-editing, from Translation Studies to Machine Learning <i>Félix do Carmo & Belinda Maia</i>	126
Machine Translation: a Secret Lover Turned into the Life Partner <i>Diego Bartolomé</i>	153
Methodology of MT Post-Editors training. Practical experience with PEMT as a part of Translation Studies at Constantine the Philosopher University in Nitra <i>Jakub Absolon</i>	170
El traductor como usuario avanzado de informática: habilidades para la traducción automática <i>Amparo Alcina</i>	194
Exploitation of a comparable corpus of user-generated content for post-editing strategies <i>Miguel Ángel Candel-Mora</i>	217
Machine translation in Spanish LSPs <i>Olga Torres-Hostench, Marisa Presas, Pilar Cid Leal</i>	265
On the correctness of machine translation: an English-Italian case study <i>Giovanni Cerasani</i>	280
Los errores de la traducción automática chino > español <i>Hong Zhang</i>	301

La postedición como nueva especialización transdisciplinar en la era de la información globalizada <i>Coral Díez</i>	332
Traducció automàtica i llengües minoritzades: el cas del sard <i>Gianfranco Fronteddu, Adrià Martín-Mor</i>	363
Workshops	390
<i>[Workshops materials are not included in this document]</i>	
When my (translation) memory is not enough: A practical workshop on how to create translation memories from the web and use them together with machine translation <i>Gema Ramírez (Prompsit)</i>	391
The tauyou Machine Translation platform <i>Diego Bartolomé (tauyou)</i>	392
MTradumàtica <i>Ramon Piqué and Adrià Martín Mor (Tradumàtica, UAB)</i>	423
Production-ready Machine Translation engines with KantanMT <i>Laura Casanellas (KantanMT)</i>	430
Post-Editing in a feedback friendly MT environment <i>Blanca Vidal (Lucy Software)</i>	432
T3L pictures	433
T3L videos	443

PLENARY SESSIONS

NEW USES OF MACHINE TRANSLATION IN THE TRANSLATION WORKSTATION

Mikel L. Forcada,

Universitat d'Alacant

<http://www.dlsi.ua.es/~mlf/>

Machine translation (MT) has now been with us for a long time. Over the years, translators have experimented with MT, and many have adopted it as a resource for their work. The incorporation of MT into the professional translation environment has depended largely on the possibilities offered by the various translation editing and management (TEM) systems, also known as computer-assisted translation (CAT) tools. The speaker will present emerging uses of MT in TEM systems, such as interactive MT (at the subsegment level), MT-based repairs of partial translation memory matches and estimates of the quality of MT at the word level in partial matches.

New uses of machine translation in the translation workstation

Mikel L. Forcada^{1,2}

¹Departament de Llenguatges i Sistemes Informàtics,
Universitat d'Alacant, E-03071 Alacant

²Prompsit Language Engineering, S.L.,
Edifici Quorum III, Av. Universitat s/n, E-03202 Elx

II International T3L Conference: Tradumatics, Translation
Technologies & Localization
10 October 2016

Outline

- 1 Machine translation everywhere
- 2 General-purpose vs. special-purpose MT
- 3 Machine translation technologies
- 4 MT in computer-aided translation: traditional way
 - Post-editing: an additional skill set
 - MT looks bad because it is always shown!
 - Post-editing: attitudes and feelings
 - Post-editing: the facts (measurements)
- 5 MT in computer-aided translation: new uses
 - Interactive/predictive machine translation
 - Target-word confidence in CAT: edit hints
 - One step further: fuzzy-match repair
- 6 Concluding remarks

Outline

- 1 Machine translation everywhere
- 2 General-purpose vs. special-purpose MT
- 3 Machine translation technologies
- 4 MT in computer-aided translation: traditional way
 - Post-editing: an additional skill set
 - MT looks bad because it is always shown!
 - Post-editing: attitudes and feelings
 - Post-editing: the facts (measurements)
- 5 MT in computer-aided translation: new uses
 - Interactive/predictive machine translation
 - Target-word confidence in CAT: edit hints
 - One step further: fuzzy-match repair
- 6 Concluding remarks

Machine translation everywhere/1

Machine translation as a technology...

- ...is not only available to professional translators to do their *magic*...
- ...but also to every *muggle* with an Internet connection!

Machine translation everywhere/2

Two main uses for machine translation:

Dissemination: **professional translators** may use it as a starting point to produce translations to be published.

Assimilation (also *gisting*): **anyone** can use MT *as is* to make sense of text in a foreign language.

F.J. Och, Google, **2012**:

In a given day we translate roughly as much text as you'd find in 1 million books. To put it in another way: what all the professional human translators in the world translate in a year, our system translates in roughly a single day.

Machine translation everywhere/3

But we will of course talk about **dissemination** applications.

Outline

- 1 Machine translation everywhere
- 2 General-purpose vs. special-purpose MT**
- 3 Machine translation technologies
- 4 MT in computer-aided translation: traditional way
 - Post-editing: an additional skill set
 - MT looks bad because it is always shown!
 - Post-editing: attitudes and feelings
 - Post-editing: the facts (measurements)
- 5 MT in computer-aided translation: new uses
 - Interactive/predictive machine translation
 - Target-word confidence in CAT: edit hints
 - One step further: fuzzy-match repair
- 6 Concluding remarks

General-purpose vs. special-purpose MT

General-purpose MT tries to address *at once* the needs of *everyone* (Google, Bing, etc.):

- It is usually free or almost free.
- It cannot meet the needs of every specific translation task.
- Getting better by using **co-text** (neighbouring text); still not too much *context* (outside the text).

Task-tuned MT is good at translating texts in a specific genre or about a specific subject.

- It comes at a cost.
- There is actually business in adapting MT to a particular task.

Outline

- 1 Machine translation everywhere
- 2 General-purpose vs. special-purpose MT
- 3 Machine translation technologies**
- 4 MT in computer-aided translation: traditional way
 - Post-editing: an additional skill set
 - MT looks bad because it is always shown!
 - Post-editing: attitudes and feelings
 - Post-editing: the facts (measurements)
- 5 MT in computer-aided translation: new uses
 - Interactive/predictive machine translation
 - Target-word confidence in CAT: edit hints
 - One step further: fuzzy-match repair
- 6 Concluding remarks

Machine translation technologies /1

There are two main groups of machine translation technologies:

- **Rule-based MT**, and
- **Corpus-based MT**

Machine translation technologies /2

Rule-based MT uses expert-written dictionaries and translation rules.

- Customization: experts edit dictionaries and rules
- Output is consistent but *mechanical*, lacking *fluency*

Machine translation technologies /3

Corpus-based MT learns to translate a corpus containing 100,000's or 1,000,000's of translated sentences.

- Customization: select texts representative of the task.
- Output: may be *deceivingly fluent (unfaithful)*.
- Main approaches:
 - **statistical machine translation** (1990's)
 - Uses **probabilistic** models estimated by counting events in the bilingual corpus used to train them.
 - **neural machine translation** (2010's).
 - Based on **artificial neural networks** inspired on how the human brain learns and generalizes.
- Such large corpora not be available for less-translated languages.

Outline

- 1 Machine translation everywhere
- 2 General-purpose vs. special-purpose MT
- 3 Machine translation technologies
- 4 **MT in computer-aided translation: traditional way**
 - Post-editing: an additional skill set
 - MT looks bad because it is always shown!
 - Post-editing: attitudes and feelings
 - Post-editing: the facts (measurements)
- 5 MT in computer-aided translation: new uses
 - Interactive/predictive machine translation
 - Target-word confidence in CAT: edit hints
 - One step further: fuzzy-match repair
- 6 Concluding remarks

MT in computer-aided translation: the traditional way

- That is... **post-editing**.

- └ MT in computer-aided translation: traditional way
 - └ Post-editing: an additional skill set

Post-editing: an additional skill set

- Real professional post-editors have first to be professional *translators* (already highly skilled!)
- With an additional set of skills (acquired through training or experience): they have to be capable of *making the most of MT output*:
 - Be familiar with typical *machine translationese*: errors, etc:
 - *Coches, camiones, **así como** motocicletas* → *Cars, lorries I **grabbed I eat** motorcycles.*
 - *Traía noticias de sus amigos* → *She brought news **from** her friends.*
 - Be good at editing, moving words or phrases around.
 - Always paying attention to the source text, avoiding being biased by the output.

└ MT in computer-aided translation: traditional way

└ MT looks bad because it is always shown!

MT looks bad because it is always shown!

- I have stolen this idea from Andy Way (EAMT 2016)¹
- Fuzzy matches from translation memories are perceived to be very useful because:
 - There is an easy, quick way to estimate and express quality (a source-side similarity called *fuzzy-match score*: FMS)
 - Proposals with FMS below a threshold are **never shown**.

But fuzzy matches are *not always available* and they are *not translations*.

- Machine-translation very often looks bad because:
 - There is no easy way to estimate and express its quality
 - It is therefore *always shown*, regardless of usefulness.

MT is *always available* and is the best possible *attempt to translate*.

¹Moorkens, J., Way, A. (2016) “Comparing Translator Acceptability of TM and SMT Outputs”, Baltic J. of Modern Computing 4:2, 141–151.

- └ MT in computer-aided translation: traditional way
 - └ Post-editing: attitudes and feelings

Post-editing: attitudes /1

Translators often find themselves *uncomfortable* or *annoyed* when postediting.

- Some feel that they should not be forced to repair the mess that a machine has made.
- Many may feel it takes longer to *fix* the output (to the same standards) than translating from scratch.
- Many hate correcting the same errors over and over again.
- Some feel they are not able to make do with MT output.

- └ MT in computer-aided translation: traditional way
- └ Post-editing: attitudes and feelings

Post-editing: attitudes /2

Some experts (e.g. Jost Zetsche) contend that *translators would feel better and be more productive* if MT was used to help them in different ways:

- Zetsche actually suggests two out of the three I will cover: *interactive/predictive MT* and *fuzzy-match repair*.²

²Zetsche, J. (2014) “The translator must always be the boss”, The Tool Box Journal, Issue 14-6-237, pp. 2–5.

Post-editing: the facts (measurements) /1

A recent study by Plitt and Masselot at Autodesk (translating with a well-tuned MT system):³

- Translation professionals who believed that postediting was slower than translating from scratch were actually translating faster!



³<http://langtech.autodesk.com/productivity.html>

- └ MT in computer-aided translation: traditional way
- └ Post-editing: the facts (measurements)

Post-editing: the facts (measurements) /2

The lesson:

- **perception of productivity** is not the same as **productivity**!
- Always **measure** yourself before making decisions!

But anyway...

Outline

- 1 Machine translation everywhere
- 2 General-purpose vs. special-purpose MT
- 3 Machine translation technologies
- 4 MT in computer-aided translation: traditional way
 - Post-editing: an additional skill set
 - MT looks bad because it is always shown!
 - Post-editing: attitudes and feelings
 - Post-editing: the facts (measurements)
- 5 MT in computer-aided translation: new uses**
 - Interactive/predictive machine translation
 - Target-word confidence in CAT: edit hints
 - One step further: fuzzy-match repair
- 6 Concluding remarks

MT in computer-aided translation: new uses

I will cover **three** possible alternative ways to use MT in the *translation editing and management* (TEM) or *computer-aided translation* (CAT) environment:

- Interactive/predictive MT
- Marking of target-word confidence in CAT proposals
- Fuzzy-match repair

Note: there are no conclusive productivity studies on these yet, no matter how *good* they *feel*.

Interactive/predictive machine translation /1

(also called *target-text mediated MT*)

- Not so new: the idea is almost 20 years old!!
- As the professional types the translation...
- ... the MT system proposes completions which are compatible with what they have typed, ...
- ... and the translator either selects one or goes on typing.

Interactive/predictive machine translation /2

- **Examples:** TransType, TransType2; a similar one is CAITRA.
- All of them use *statistical MT*, but could use *rule-based MT*
- Commercially available: *AutoSuggest* in SDL Trados, *AutoComplete* in Wordfast Classic, or *AutoWrite* in Déjà Vu X3.
- Experimental free/open-source plug-in for OmegaT and Apertium by Daniel Torregrosa-Rivero.⁴

⁴<https://github.com/dtr5/apertium-cli-omegat>

- └ MT in computer-aided translation: new uses
 - └ Interactive/predictive machine translation

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: C

- └ MT in computer-aided translation: new uses
 - └ Interactive/predictive machine translation

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: Ce

- └ MT in computer-aided translation: new uses
 - └ Interactive/predictive machine translation

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: Ce projet de loi

- └ MT in computer-aided translation: new uses
 - └ Interactive/predictive machine translation

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: Ce projet de loi est

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: Ce projet de loi est très semblable

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: Ce projet de loi est très semblable a

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: Ce projet de loi est très semblable au

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: Ce projet de loi est très semblable au projet de loi

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: Ce projet de loi est très semblable au projet de loi q

Interactive/predictive machine translation /3

Source: This bill is very similar to its companion bill which we dealt with yesterday.

Typing the target: Ce projet de loi est très semblable au
projet de loi que ...

Target-word confidence: *edit hints* /1

- When CAT systems bring up a proposal for a new segment. . .
- . . . they highlight the *source words* that do not match.

New sentence: Connect the **printer** to the computer

Source from TM: Connect the **scanner** to the computer

Target from TM: Connecteu l'escàner a l'ordinador.

Target-word confidence: *edit hints* /2

Wouldn't it be **cool** for the tool to also highlight the *target words* that **need to be changed**?

New sentence: Connect the **printer** to the computer

Source from TM: Connect the **scanner** to the computer

Target from TM: Connecteu **l'escàner** a l'ordinador.

Target-word confidence: *edit hints* /3

Can you use MT to do that?

- 1 determine the words that changed in the source side
 - Already done by the CAT tool
- 2 decide which parts of the target side of the proposal correspond to changed parts of the source side of the proposal
 - We will use MT for that
- 3 mark them.

Target-word confidence: *edit hints* /4

New sentence: Connect the **printer** to the computer

Source from TM: Connect the **scanner** to the computer

Target from TM: Connecteu l'escàner a l'ordinador.

Get **clippings** around the source-side mismatches, with *context*:

- 1 “the **scanner**” → “the **printer**”
- 2 “the **scanner** to” → “the **printer** to”
- 3 “**scanner** to” → “**printer** to”
- 4 “connect the **scanner**” → “connect the **printer**”
- 5 “**scanner** to the computer” → “**printer** to the computer”
- 6 . . .

Target-word confidence: *edit hints* /5

Use MT to determine which parts of the target correspond to mismatched parts of the source:

New sentence: Connect the **printer** to the computer

Source from TM: Connect the **scanner** to the computer

Target from TM: Connecteu l'escàner a l'ordinador.

Machine-translate clippings at source-side mismatches and match them in the target side of the proposal.

- 1 “the scanner” [2,3] → “l'escàner” [2,3]
- 2 “the scanner to” [2,4] → “l'escàner a” [2,4]
- 3 “scanner to” [3,4] → “escàner a” [3,4]
- 4 “connect the scanner” [1,3] → “connecteu l'escàner” [1,3]
- 5 “scanner to the computer” [3,6] → “escàner a l'ordinador” [3,6]

All match the target! (This may not always be the case).

Target-word confidence: *edit hints* /6

The matched target words are candidates to being edited:

New sentence: Connect the **printer** to the computer

Source from TM: Connect the **scanner** to the computer

Target from TM: Connecteu l'escàner a l'ordinador.

Target-word confidence: *edit hints* /7

At each target word, evidence from different clipping lengths is *weighted* (using a pre-learned set of weights) and *thresholded*...

New sentence: Connect the **printer** to the computer

Source from TM: Connect the **scanner** to the computer

Target from TM: Connecteu **l'escàner** a l'ordinador.

... to mark the words to be edited.

It could also be used to mark the words to keep.

Target-word confidence: *edit hints* /8

- Note that **no single machine-translated word is shown** to the translator.
- MT is **only used where it agrees with the translation memory**
- **Any kind of translation source** can be used (technology-agnostic).
- Method extensively studied in Miquel Esplà-Gomis's PhD thesis.
- As far as I know, this has not been implemented commercially.
 - But there is a free plug-in for OmegaT.⁵
- One limitation: does not mark where inserts are needed (yet!).

⁵<http://www.dlsi.ua.es/~mespla/edithints.html>

Target-word confidence: *edit hints* /9

The screenshot shows the OmegaT-3.0.7 :: lat6 interface. The main window is titled "Editor - ci-ny-es.odt". The document content is in Spanish and includes the text "Ciprofloxacina Nycomed y nombres asociados" and "Actira and associated names". The text is highlighted in green and red. The document also contains a section titled "INFORMACIÓN BÁSICA" and a paragraph describing the antibiotic Ciprofloxacina.

On the right side, there is a "Fuzzy Matches" panel. It lists two matches:

- 1) Actira y los nombres asociados
Actira and associated names
<50/40/54%
/home/miquel/Dropbox/exempleprojecteOmegaT/lat6
- 2) Avelox y los nombres asociados
Avelox and associated names
/home/miquel/Dropbox/exempleprojecteOmegaT/lat6

Below the fuzzy matches is a "Glossary" panel, which is currently empty.

At the bottom of the interface, there are tabs for "Dictionary", "Machine Translation", "Multiple Translations", "Notes", and "Comments". The "Machine Translation" tab is selected. The bottom right corner shows the progress "0/36 (0/36, 36)" and "41/30".

One step further: fuzzy-match repair /1

The above procedure can be taken one step further to **repair** the target proposal.

Machine-translate what changed from the proposal to the new sentence.

Then pair

- translations of new-sentence clippings to
- translations of proposal clippings

to build “repair operators”.

One step further: fuzzy-match repair /2

Translation of proposal clipping → Translation of source mismatch clipping.

- 1 “l’escàner” [2,3] → “la impressora”
- 2 “l’escàner **a**” [2,4] → “la impressora **a**”
- 3 “escàner a” [3,4] → “impressora per”
- 4 “**connecteu** l’escàner” [1,3] → “**connecteu** la impressora”
- 5 “escàner **a l’ordinador**” [3,6] → “impressora **a l’ordinador**”

Overlap in **bold**: overlap with the proposal is desirable.

One step further: fuzzy-match repair /3

Change the corresponding parts in the target proposal to obtain one or more approximate **repaired** translations

Raw proposal = “Connecteu l’escàner a l’ordinador”

Repair 1: “Connecteu **la impressora** a l’ordinador”

Repair 2: “Connecteu **la impressora a** l’ordinador”

Repair 3: “Connecteu l’**impressora per** l’ordinador”

Repair 4: “**Connecteu la impressora** a l’ordinador”

Repair 5: “Connecteu l’**impressora a l’ordinador**”

One step further: fuzzy-match repair /4

Some repairs are not adequate. Which are the best repairs?

- Not having easily-spotted language errors (*l'impressora*)
- Those containing a wider source-side context?
- Those having target-side context on both sides of the actual change?

Use **quality estimators**, machine-learned from reference translations, to predict the most promising repairs.

One step further: fuzzy-match repair /5

- Fuzzy-repair operators rely on MT (danger!)
- But the history of matching of clippings in previous translation units can be used as a quality feature in future repairs:
 - the MT system and the translation memory “say the same”
→ more confidence
- Similar functionality commercially available:
 - Déjà Vu X3 DeepMiner
 - MemoQ
- Free plug-in for OmegaT will soon be released (by John E. Ortega and Miquel Esplà-Gomis).

Outline

- 1 Machine translation everywhere
- 2 General-purpose vs. special-purpose MT
- 3 Machine translation technologies
- 4 MT in computer-aided translation: traditional way
 - Post-editing: an additional skill set
 - MT looks bad because it is always shown!
 - Post-editing: attitudes and feelings
 - Post-editing: the facts (measurements)
- 5 MT in computer-aided translation: new uses
 - Interactive/predictive machine translation
 - Target-word confidence in CAT: edit hints
 - One step further: fuzzy-match repair
- 6 Concluding remarks**

Concluding remarks

- Post-editing is not the only way to use MT in the translator's workstation.
 - But professional attitudes towards post-editing do not always agree with measurements
 - Perception of productivity is not the same as productivity: **if in doubt, time yourselves.**
- But MT may also be used to:
 - Interactively help translators as they type a new translation.
 - Marking words to be edited in a proposal from a translation memory.
 - Actually repairing a proposal from translation memory.
- No conclusive data yet on actual productivity gains for alternative uses of MT.

Acknowledgements and thanks

Acknowledgements:

- People in the Transducens research group: Miquel Esplà-Gomis, John E. Ortega, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez Daniel Torregrosa-Rivero.
- Funding from the Spanish and the Kazakh governments for research and plugin development.

I thank...

- ... the organizers for inviting me...
- ... and all of you for your attention!

These slides are free

This work may be distributed under the terms of either

- the Creative Commons Attribution–Share Alike licence:

`http:`

`//creativecommons.org/licenses/by-sa/4.0/`

- the GNU GPL v. 3.0 Licence:

`http://www.gnu.org/licenses/gpl.html`

Dual license! E-mail me to get the sources: `mlf@ua.es`

PANORAMA 2020: WHICH WILL BE THE MOST SOUGHT-AFTER PROFESSIONAL PROFILES IN THE TRANSLATION MARKETPLACE?

Olga Blasco

The Rosetta Foundation

<https://www.therosettafoundation.org/>

In response to the global paradigm shift over the last decade, there have been rapid technological changes to achieve total connectivity, maximum automation, high productivity and non-stop service to clients. Small, medium and large translation companies are adopting innovative commercial and operational strategies to position themselves in an increasingly competitive market. What professional profiles do they need to make it possible now and in the future, and why?

Panorama 2020:
professional profiles
in demand
for the translation
marketplace

Olga Blasco – Consultant
Business Strategy for Growth



What we will cover



The
translation
industry
today

Human Talent
- Challenges

Professional
profiles
- 2020

My expertise and focus is on... (Short Version)

Business growth strategy

Organizational development

Solving difficult problems

(And the long version...!)

Leadership & Coaching

Organizational Development

Multi-cultural Team Building

Operations Optimization

Networking & Partnerships

Financial Planning & Budgeting

Mergers & Acquisitions

Talent Sourcing & Development

Due Diligence & Research

Business Analytics

Communications Strategy

Negotiation & Mediation


Business Development

Event & Campaign Management

My main focus is:

- driving the strategic direction of the organization into the next phase
- financial sustainability,
- partnerships

we connect
11,000+
volunteer translators,
through our **Trommons crowdsourcing** platform,
with over 250 **non-profit** organizations all over the world



The
ROSETTA
Foundation

TRANSLATION CHANGES LIVES - JOIN US AS A VOLUNTEER - BECOME OUR DONOR.
#Languagematters

WWW.THEROSETTAFOUNDATION.ORG

The translation industry today












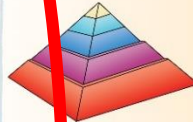
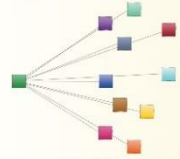
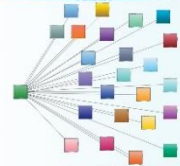
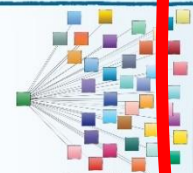
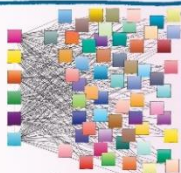

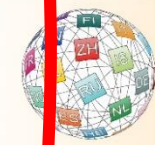






Size of the industry

The industry is diverse and technology-driven, with an increasing impact on both global and regional economies. Here are three key facts to consider:

- The size of the overall global language industry in 2016 is estimated at \$40 Billion (USD), with estimates of up to \$45 Billion by 2020
- The projected growth rate is 6.5-7.5% annually through 2018
- The size of the language *technology* industry is estimated at \$29 Billion

Source: Common Sense Advisory



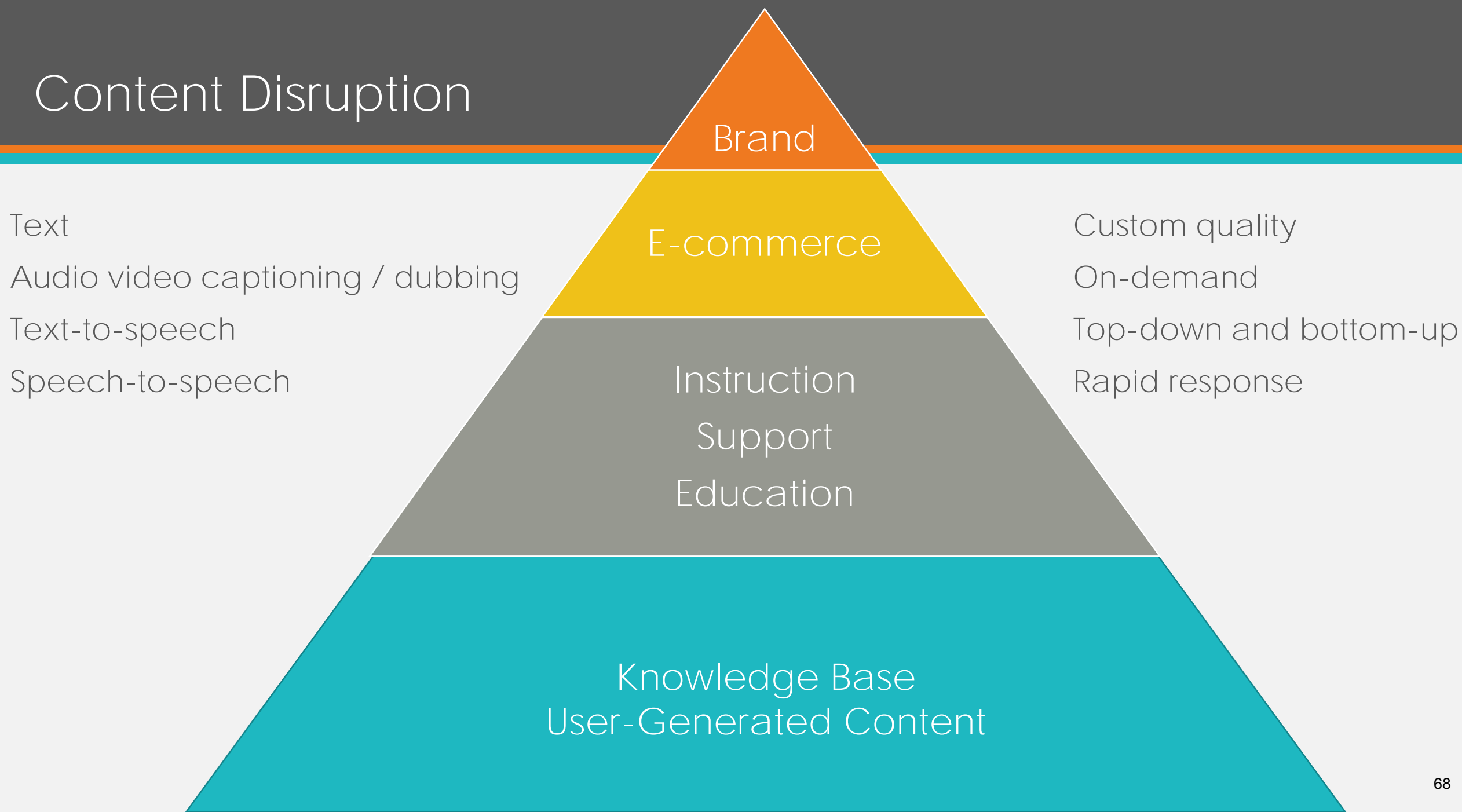
	1980 TRANSLATION	1990 LOCALIZATION	2000 GLOBALIZATION	2010 INTEGRATION	2020 CONVERGENCE	2030 SINGULARITY
INDUSTRY FOCUS	 Documents	+  Software	+  Simship	+  Integration in enterprise systems	+  Embedded in every app, on every screen	+  Ubiquitous
TYPE OF CONTENT	 Paper	+  Digital	+  Static Web	+  Dynamic Web	+  Personalized	+  Unlimited
LANGUAGES	 $1 > 10$	 $1 > 25$	 $1 > 40$	 $6 <> 60$	 $60 <> 60$	 $100 <> 150$
TECHNOLOGY	None	TM and terminology software	Workflow (GMS)	MT and Advanced Leveraging	Real-time customized MT	FAUT
DATA	Glossaries	Project TMs	Centralized TMs	Limited shared data	Web crawled	Plug and Play
COMMUNICATIONS	 B2B G2C B2C	+  C2B	+  C2C	+  Social	+  Internet of Things	+  M2M

Evolution of the Translation Technology Landscape

Source:



Content Disruption



Common Friction Points...

- Processing and managing files (end to end)
- Large amount of administration and data entry for PMs
- Getting the correct resources in time to get the job done
- Collect and report data for performance measurement
- Spend control
- Recruitment and management of external resources
- Managing the **customers' expectations** (SLAs)

Too many moving parts, intrinsically labour-intensive, and human error-prone.



Technology, Technology, Technology

To remain competitive everyone is looking for:

- cloud-based simplicity,
- seamless interoperability
- automation



Most-wanted solutions

So the focus now is on:

- UX (user experience) design = easy-to-use systems
- comprehensive range of APIs
- robust TMS with accounting connectivity
- talent matching functionality - **“find the needle in the hay stack”**
- MT to maximize value through high productivity and custom quality outputs.
- Translation collaborative environment = in-context, queries, terminology, QA
- Dashboards - analytics / modelling = performance, forecasting and capacity



Industry shake-up

Disruptive start-ups

Open Source

Multi-coloured supply chain

MT as the new normal

Real-time

Investment in NMT and Artificial Intelligence R&D

New pricing models

Petabytes of data

1 PB = 1,000,000,000,000,000 B = 10^{15} bytes = 1,000 terabytes

Mergers and Acquisitions



Mergers and Acquisitions is a Growing Trend

Juicy recipe of diversification by acquisition:

Buy clients + brands + technology + expertise =
gain market share fast + beat profit compression +
accelerate return on investment

Ideal company valuable for acquisition:

- >USD 4 million revenue and 20% EBITDA
(earnings before interest, tax, depreciation, and amortization)
- client loyalty in one or two regulated industries
- solid brand and niche expertise

Buying Was the Easy Part: M&A Integration in the Cold Light of Day

by Olga Blasco on July 29, 2016

Features



We know the translation market is growing and with it the number of mergers and acquisitions. What we do not know yet is the true impact all this activity will have on the overall market—clients, technology companies (established or disruptive), and the supply chain.

If we take a look at the M&A stage, investor-backed and publicly traded companies are set to follow the juicy recipe of diversification by acquisition; buying new clients, brands, technology, and expertise in order to gain market share fast, beat margin compression, and accelerate returns.

My article on 29 July 2016



M&A greenfield



The vast majority of LSPs are <\$1 million
There are hundreds of companies getting ready for exit = sale

Why M&A is relevant to today's topic

Market – “**David vs. Goliath**” environment

- Big corporations = economies of scale
 - Small companies = agile and flexible
- = > “If you **can't** beat a competitor, buy it”

Technology – “**to integrate or not to integrate**”

- maintain multiple systems and connect as needed
 - build homegrown *über-platform*
- = > The choice has consequences

Talent – “**new** company, same people?”

- risk-conscious clients want disintermediation, credentials and scale
 - adoption of new technologies is still a hard-sell rabbit-out-the-hat innovation and business-as-usual
- = > “**Don't** throw the baby out with the bath water”

Human Talent - Challenges

Look at the world from a different lens....



The photographer's profile

In 2013 Nara was 7 years old and was living in Bangkok.

Her mother Monica is a seasoned amateur photographer.

Nara had no previous hands-on photography experience, except **messing with her mother's smart** phone.

She was given a Lumix point and shoot **camera to experiment one afternoon...**

Maybe she was a fast **learner, a lucky rookie or naturally talented, but...**

... years ago this would not have been possible.

How companies view automation

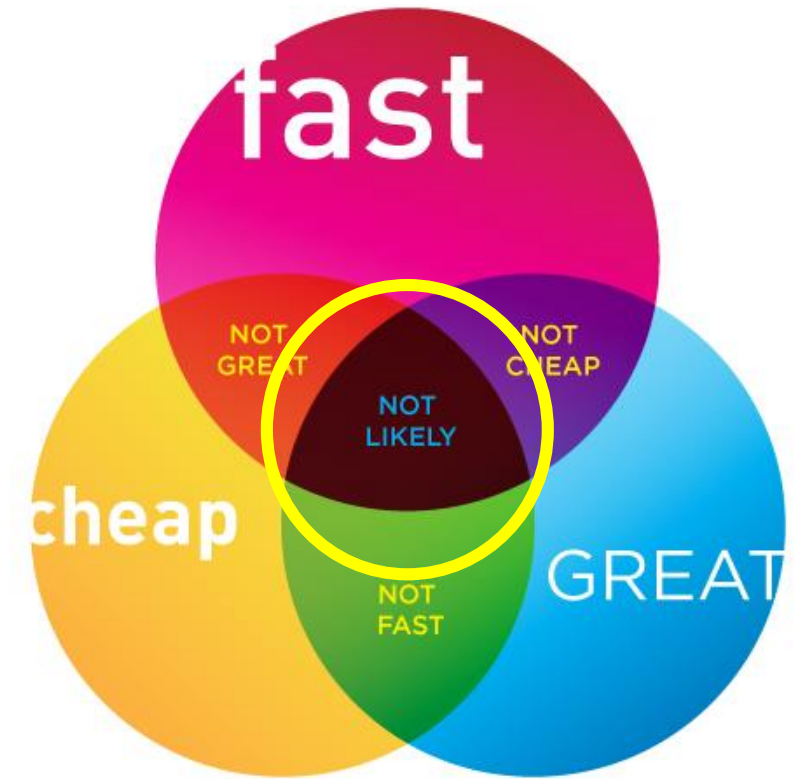
A productivity booster to **do more with less, faster**

A way to **simplify logistics beyond traditional roles and skill sets**

A **substitute for experience** that allows to **“cast a wider net”** on resources available:

- Shorten ramp-up and learning curve to increase capacity
- Make road testing and attrition manageable

Ultimately, **a means to reduce the cost of doing business.**



© copyright 2011 | Compass Creative Studio Inc. compasscreative.ca

Top Challenges

Thanks to technology, we can do things we didn't think possible even 10 years ago.

But still today the fact remains that without the **right human resources** available at the **right time**, it is impossible to deliver language services.

Risk-conscious clients want **credentials**.

The real issue to ensure that the **chosen and qualified capacity** will **available** there when it is needed, and can scale as required at less total cost.

Aggressive cost reduction without the proper **due diligence on viability** can end in disaster – this happens still far too often.



Who Are
you?



Sourcing Translators – Balancing act

Transcreators, client specialists, domain specialists,
casual generalists, post-editors....

Experience – Flexibility

Mature vs New

- train experienced resources (who are largely set in their ways)
- accelerate the learning curve of newcomers (under the right controls)

Price – Value Add

Small LSPs vs Freelancers

- Small LSPs = expert resource-broker platforms to support scale
- Unless value proposition is clear, even big organizations choose individuals

Loyalty - Capacity

Task-allocation portals vs Enthusiast-crowdsourced communities

- Motivation is not only about compensation but keeping people engaged as they work harder and better.
- Do individuals guarantee capacity when they are a code number dealing with a task allocation portal?
- Can you expect loyalty without some form of appreciation?



Roles in the business process

The business model that generates the most sales is a **client-centric** one:

- organized around **business units** (geographic, industries, hybrid)
- with a **shared services** team (VM, HR, Finance) that supports all

In the era of **“convergence”** it's time to ask questions such as:

- What is the **rationale behind each role**, what makes it necessary and unique in its value add to a book of business?
- How do you resource a book of business that is tech-enabled where you **manage by exception**?
- What is required operationally to offer **custom quality solutions**?
- **How many trainers and “implementers”** are needed to get tech-enabled?



Organizational structures

Here are examples of **overlaps** in the traditional demarcations:

Account Manager (Sales) and **Program Manager** (Ops)
– both have direct relationship with the client

Solutions Architect (systems strategy) and **Engineer** (doer)
– both solve technical issues

Project Manager (project and budget owner) and **Quality Manager** (quality owner)
– both interact with clients and vendors

Quality Manager (quality owner) and **Vendor Manager** (capacity & spend owner)
– both interact with vendors

Time to **challenge the status quo** in relation to roles, hierarchies, default demarcations and duplications.



Back to Basics 2.0?

These traditional demarcations respond to skill set but also the workload traditionally involved in each piece.

In a world of “quality at speed”, these demarcations create too many human touchpoints, even when technology is implemented.

Ideally one would want to:

- Enable human scale with the use of technology
- “**Merge**” some of those roles
- Focus on the right combination of technical expertise and soft skills to match people and roles



Professional profiles - 2020

Machines

Drones

VR (Virtual Reality)

AR (Augmented Reality)

Google Home, Amazon Echo

Driverless cars

3D Printing

Industrial and Medical Robots

...and yes, Artificial Intelligence

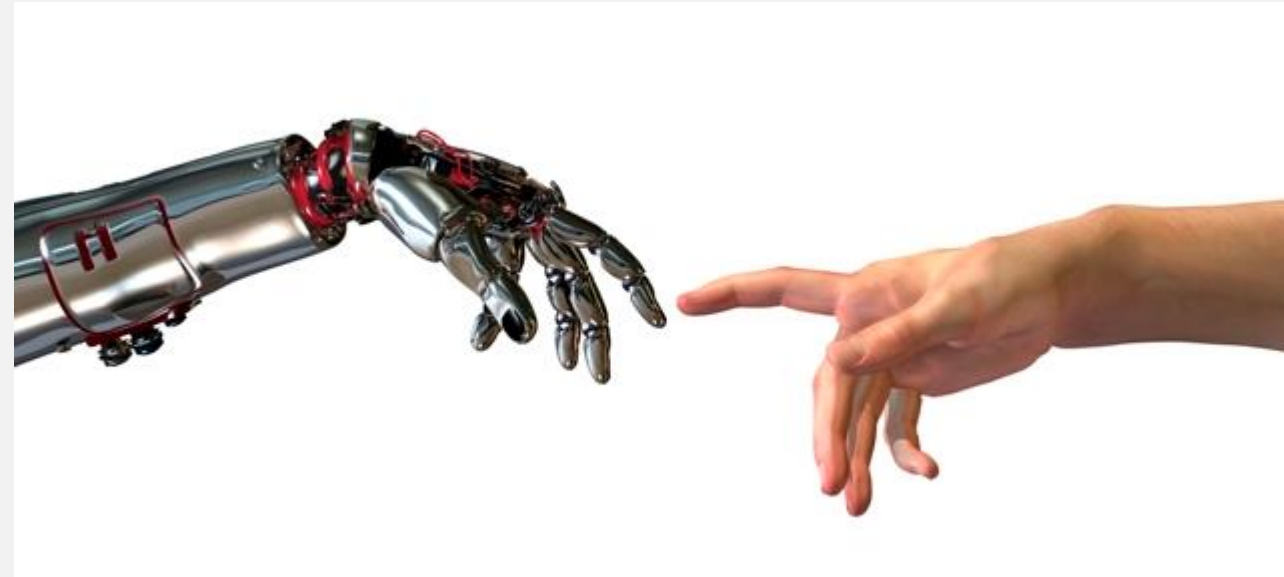


Artificial Intelligence

Google Neural Translate

*“**Bridging** the Gap Between Human and Machine **Translation**”?*

- Real breakthroughs
- Too much training overhead
- Still a long way from HT
- Incremental improvements



But we are heading towards a **Value (R)evolution**

Adapt or Die

Net-destroyer or net-creator?

Both.

Theory of Instrumental Convergence

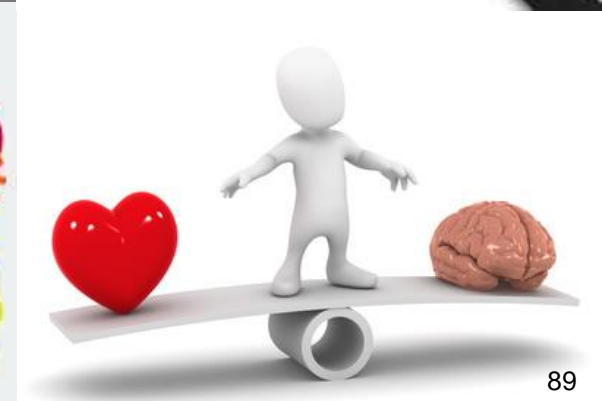
What will Artificial Intelligence R&D struggle to replicate?



Human Value: adaptability, flexibility, and resourcefulness

Soft skills that are and will be valued in business:

- Emotional Intelligence
- Creative Thinking
- Customer-centric Problem-solving
- Strategic Decision-making
- Human-Centered Design:
Inspiration-Ideation-Implementation
- (Virtual) Team Work & Team building
- Circles of Influence
- Awareness of Cultural Differences



Professional profiles in demand - 2020

Translators

- Language consultants
- The end of writing
- Real-time, cloud-based technology interaction
- Highly productive

Professional profiles in demand - 2020

Technology Product Managers

- Human-Centered Design
- UX (User Experience)
- Understand the “needs” **and** “pains”

Professional profiles in demand - 2020

Program Managers

- High EQ
- Big picture
- Understands the business
- Sense of urgency

Professional profiles in demand - 2020

Data Analysts

- “*Moneyball*” (movie)
- Use the right parameters
- Understand indicators
- Modelling

Professional profiles in demand - 2020

Talent Managers

- **Capacity “rainmakers”**
- High EQ
- Resourceful and creative
- Resilient – tough job (even with tech-enabled sourcing)

Professional profiles in demand - 2020

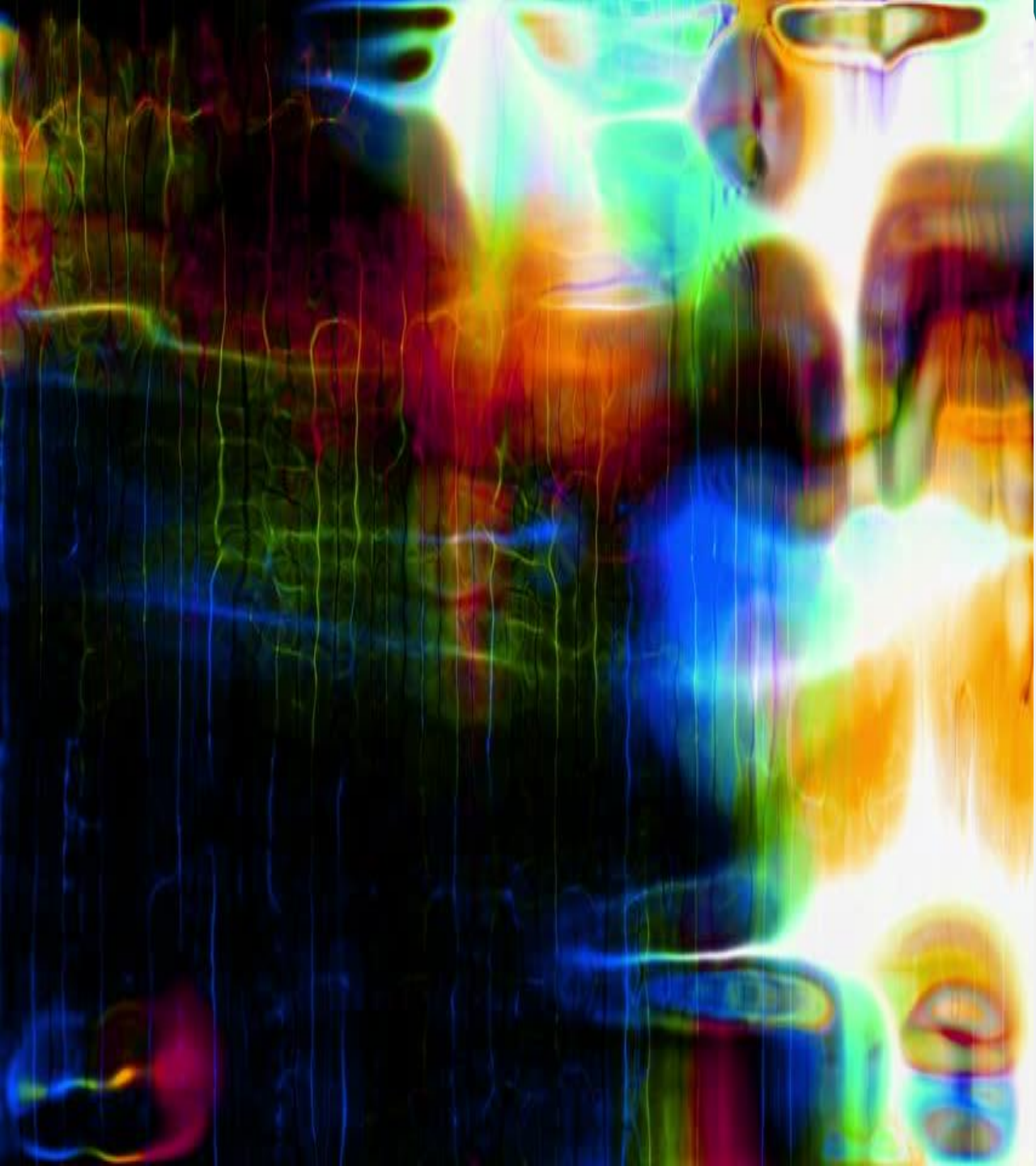
Trainers & Implementers

- Speak plain
- Empathetic
- Enthusiastic but mindful of strengths and limitations

Professional profiles in demand - 2020

And of course...

- Computational Linguists
- Developers
- Engineers



CONFERENCES

THE QUALITY OF POST-EDITED MACHINE TRANSLATION: WHAT'S IT TO YOU?

Van Egdom Gys-Walt

Zuyd University of Applied Sciences

<https://www.peterlang.com/view/product/23511>

Machine translation and post-editing are often subjected to stern and lofty criticism. The community of translators seem to speak with a unified voice when airing the complaint that, with time, the profession will be endangered or, at least, radically transformed by both phenomena, limiting her/his role to that of a post-editor or even to that of an all-round linguistic or intercultural expert, and that, as a consequence of this transformation, quality standards will no longer be upheld.

Instead of easing or allaying the fears of the community, translation studies scholars have contributed to those fears by comparing the quality of post-edited machine translations to target texts made 'from scratch'. Although the qualitative superiority is established or corroborated time and again, these studies alert us to the fact that the loss of quality is probably offset by other factors (e.g. an increase in productivity)

Beyond the theoretical ken of the aforementioned studies lies a terrain that is fraught with perilous obstacles, the 'hidden' domain of quality. The assumption that quality is a stable concept or that quality is only defined by the profession has left many scholars us mired in error for decades. These kneejerk assumptions are reflected in the aforementioned studies. In a recent study, we have felt the need to venture a leap into the great unknown by defining quality as perceived quality (per definitionem).

In our practice-oriented research, the students of the Zuyd simulated translation bureau were asked to produce four PE-versions of two English source texts. Each version was related to a degree of editing. The eight post-edited machine translations were distributed to LSP's as well as to end-users, along with a survey form. With our pilot-study, we have hoped to gain demonstrably deeper understanding of the quality standards of both respondent groups, with a view to developing a rationale for what is commonly called the 'fitness for purpose' of the PE text.

The Quality of Post-edited Machine Translation

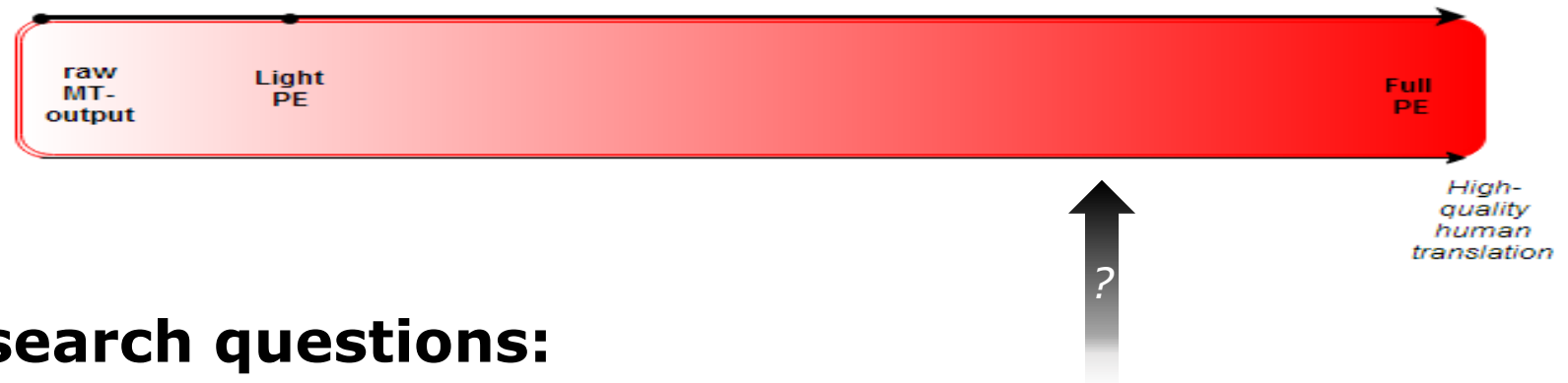
what's it to you?

Gys-Walt van Egdom (PhD)
Zuyd University of Applied Sciences
Vrije Universiteit Brussel
T3L Conference, Barcelona
10 October 2016

Background and Research questions

Point of departure:

- PE guidelines make no mention of end users;
- PE guidelines offer no alternative to light PE and full PE



Research questions:

- What is the effect of an increasing PE effort on text quality as perceived by TSPs and end users?
- What is the effect of an increasing PE effort on the image of the text producer as perceived by TSPs and end users?

Design:

- Two experiments with PE texts:
 - └ TSPs (N = 89): 'Protecting you';
 - └ end users (N = 77): 'Protecting you';
- Between-subjects manipulation of PE efforts*
- Questionnaire with measure scales for quality perception:
 - (1) content (4) style
 - (2) language use (5) usability
 - (3) logic (6) image of the text producer
- Separate MANOVAs for TSPs and end users
- Repeated contrasts

and

- Calculation of edit distance

Manipulation of PE effort

V1 – minimal PE:

instructions: correct names, maintain anaphoric relation, parse long sentences, avoid ambiguity

≈ pre-edited MT

V2 – light PE:

instructions: [V1+] correct grammatical and lexical errors (inversions, congruence and juxtapositions)

V3 – moderate PE:

instructions: [V1+V2+] improve logic, create cohesion and correct terminology

V4 – full PE:

instructions: [V1+V2+V3+] improve style and add idiomatic constructions

Loosely based on Mossop (2001/2014)

MANOVA results: 'Protecting you'

TSPs	Light vs. minimal	Moderate vs. light	Full vs. moderate
Content	$p < 0.001$	$p > 0.05$	$p > 0.05$
Language use	$p < 0.001$	$p < 0.001$	$p > 0.05$
Logic	$p < 0.001$	$p < 0.001$	$p > 0.05$
Style	$p < 0.001$	$p < 0.001$	$p > 0.05$
Usability	$p > 0.05$	$p < 0.001$	$p > 0.05$
Image	$p < 0.001$	$p < 0.001$	$p < 0.05^{**}$

** not in the expected direction

End users	Light vs. minimal	Moderate vs. light	Full vs. moderate
Content	$p < 0.001$	$p < 0.001$	$p > 0.05$
Language use	$p < 0.001$	$p < 0.001$	$p > 0.05$
Logic	$p < 0.005$	$p < 0.005$	$p > 0.05$
Style	$p < 0.001$	$p < 0.001$	$p > 0.05$
Usability	$p < 0.001$	$p < 0.001$	$p > 0.05$
Image	$p < 0.001$	$p < 0.005$	$p > 0.05$

Edit distance:
'Protecting
you'

Phishing'	Raw MT output	Minimal PE	Light PE	Moderate PE
Minimal PE	94,68%	XXXXXXXXXX	XXXXXXXXXX	XXXXXXXXXX
Light PE	90,43%	88,02%	XXXXXXXXXX	XXXXXXXXXX
Moderate PE	87,25%	85,01%	96,06%	XXXXXXXXXX
Full PE	70,82%	70,44%	73,57%	74,79%

Conclusions:

- Overall, PE effort has a significant positive effect on the judgements of TSPs and end users about text quality and text producer image
- An exception made for full PE
- There is a remarkably high level of agreement between TSPs and end users

Prospects and Limitations

Limitations:

- Limited text (type)
- Convenience sample
- Operationalisation of PE effort

Prospects:

- PE competence model (Van Egdom, forthcoming)
- Assessment of PE (Van Egdom et al., forthcoming)
- PE-apptitude test
- Format for PE-brief (cf. Allen, 2003)

POST-EDITING SUBOPTIMAL MACHINE TRANSLATION: WHY, WHEN AND HOW?

Nora Aramberri

University of the Basque Country (UPV/EHU)

<http://www.ehu.eus/ehusfera/ixa/>

It is well known that machine translation (MT) systems do not always provide high-quality output. This does not necessarily mean that translators cannot benefit from it coupled with post-editing. Nevertheless, little attention has been devoted to analysing what happens in these scenarios. In this line, this presentation aims to start a discussion on aspects relating to MT quality and the post-editing task, specifically, potential advantages, changes in the post-editing approach and training.

Post-editing suboptimal machine translation: why, when and how?

Nora Aranberri

IXA group – University of the Basque Country

***2nd International T3L Conference: Tradumatics, Translation
Technologies & Localisation
“Translators and machine translation”***

10-11 October 2016 - Universitat Autònoma de Barcelona

OUTLINE

- × Post-editing productivity
- × Machine translation quality
- × Suboptimal quality
- × Experience from workshop
 - + Feedback
 - + Training
- × Why, when and how?

POST-EDITING (PRODUCTIVITY)

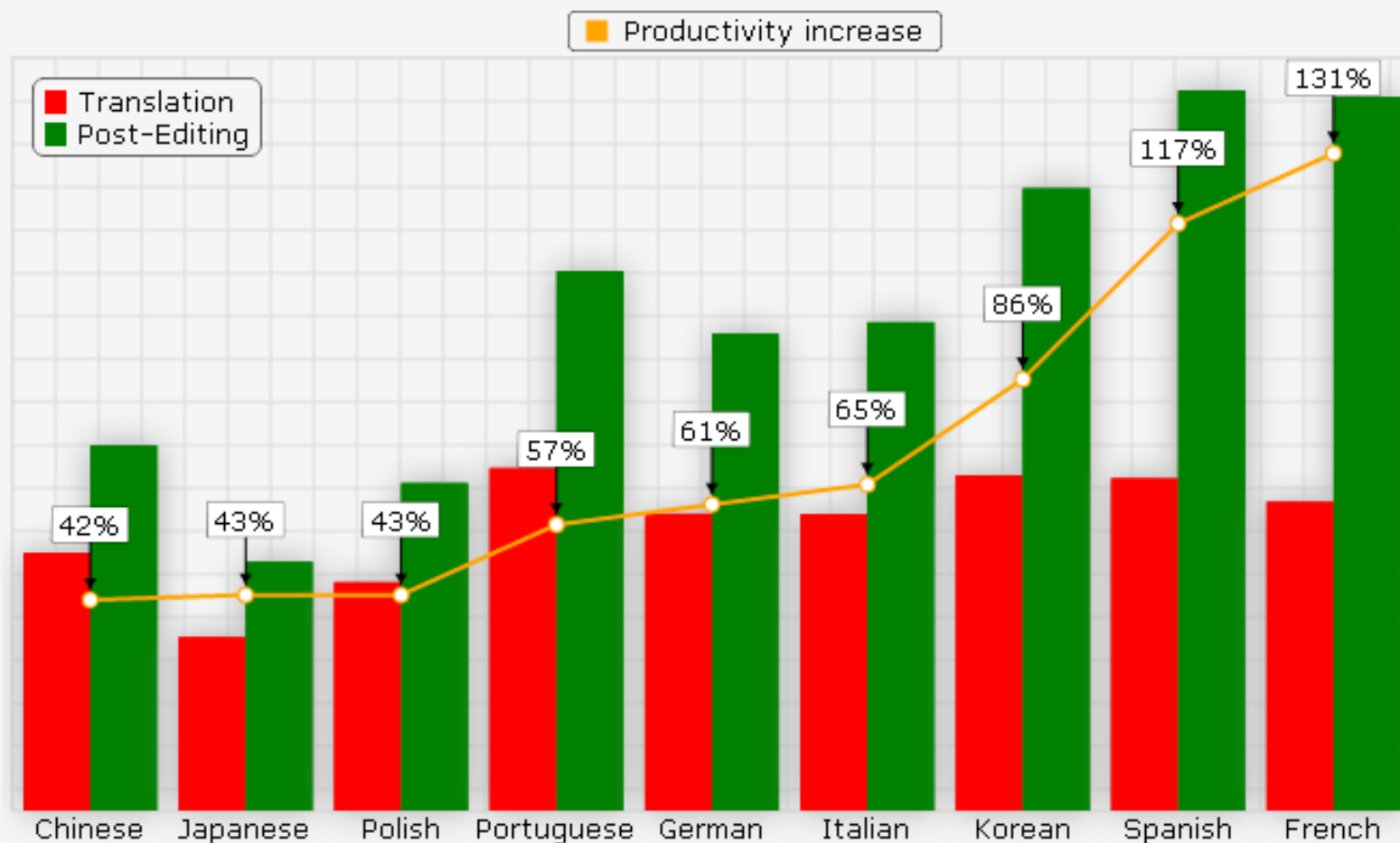


POST-EDITING PRODUCTIVITY

Autodesk (2011)

Productivity per Language – Translation vs Post-Editing

For all languages tested – in fact for all 37 test participants –, post-editing productivity was significantly higher than translation productivity.



POST-EDITING PRODUCTIVITY

“The findings suggest that translators have higher productivity and quality when using machine-translated output than when translating without it”

Guerberof Arenas, A. 2014. Correlations between productivity and quality when post-editing in a professional context. *Machine Translation*, 28: 165.

POST-EDITING EXAMPLE

× EN Source

+ You defined the possibilities and we redefined the phone.

× ES MT (Google Translate)

+ Definió las posibilidades y redefinimos el teléfono.

× ES Post-editing

+ ~~Definió~~ Tú definiste las posibilidades y nosotros redefinimos el teléfono.



× ES source

+ Tú definiste las posibilidades y nosotros redefinimos el teléfono.

× EU MT

+ Aukerak definitu zenituen eta telefonoa berdefinitu dugu.

× EU post-editing

+ Zuk aukerak definitu zenituen eta guk telefonoa berdefinitu dugu.

MACHINE TRANSLATION QUALITY

- × Autodesk:

- + Customized Moses machine translation engine

- × Gueberof:

- + Customized Moses machine translation engine
with a BLEU score of 0.60

MACHINE TRANSLATION SCENARIOS

- ✖ Customised systems vs generic systems
- ✖ Mainstream vs low-resourced languages
- ✖ MT for specific jobs vs MT for daily jobs



MACHINE TRANSLATION QUALITY

× ES source

- + Tú definiste las posibilidades y nosotros redefinimos el teléfono.

× EU MT (Basque Government – Lucy software)

- + Zuk aukerak definitu zenituen eta guk telefonoa berriz definitzen dugu.

× EU post-editing

- + Zuk aukerak definitu zenituen eta guk telefonoa berriz definitut~~zen~~ dugu.

× ES source

- + Hemos diseñado Galaxy S7 y S7 edge dando rienda suelta a nuestra imaginación.

× EU MT (Basque Government – Lucy software)

- + S7 Galaxy eta edge-a S7 diseinatu dugu aho-uhal askea gure irudimenari emanaz.

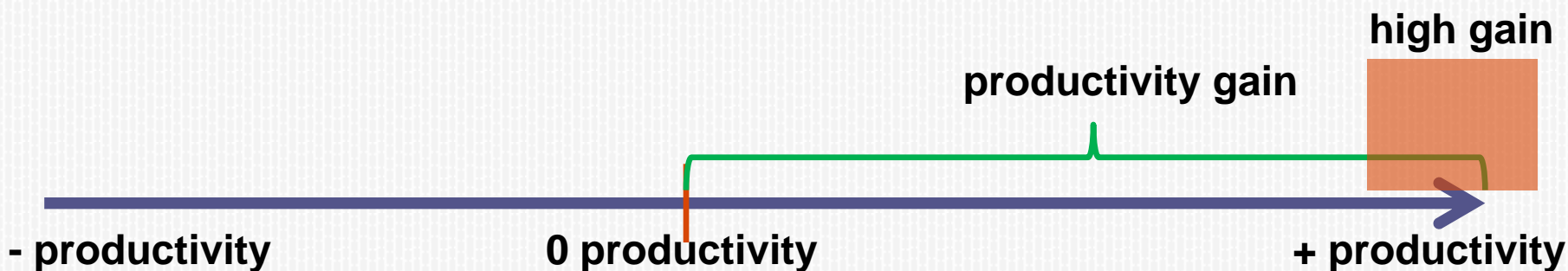
× EU post-editing

- + S7 Galaxy eta S7 ~~edge-a~~ diseinatu ditugu~~dugu~~ aho-uhal gure irudimenari ~~askatuz~~.



WHAT IS SUBOPTIMAL MT QUALITY?

- × Optimal translation: A quality that allows a translator to be **HIGHLY** productive when post-editing as compared to translating from scratch.
- × Suboptimal translation: anything below optimal quality.
 - + A quality that allows a translator to be somewhat more productive when post-editing.
 - + A quality that results in no difference in productivity.
 - + A quality that does not allow a translator to be more productive when post-editing as compared to translating from scratch.



- × Successful BLEU scores: ~40-60 points depending on the language pair
- × BLEU scores for Basque: ~5-20 points depending on test-set

EXPERIENCE FROM WORKSHOP: SET-UP

- ✗ Independent workshop for professionals (autumn 2015)
- ✗ Aim: explore Basque post-editing
 - + post-editing work
 - + provide a space for discussion
- ✗ Set-up:
 - + 10 professional translators, 8 weeks
 - + 5 face-to-face sessions
 - + Theoretical aspects: MT architectures, post-editing
 - + Practical exercises: 4 post-editing jobs + 1 productivity job per week
 - + Tools
 - ✗ Basque Government's MT system (mainly)
 - ✗ Report by Emakunde (mainly)
 - + Manual evaluation of a number of segments



ES -> EU

EXPERIENCE FROM WORKSHOP: FEEDBACK

Feedback n° 1:

“MT output quality varies considerably within a document. Regardless of the overall (low) quality, we can reuse segments or portions of segments successfully!”



EXPERIENCE FROM WORKSHOP: FEEDBACK

Feedback n° 2:

“Even low/modest quality output often includes relevant terminology and provides a syntactic structure to start with.

It helps to avoid the “blank page” syndrome.”



EXPERIENCE FROM WORKSHOP: FEEDBACK

Feedback nº 3:

“Post-editing task: from correcting (small) mistakes to identifying useful chunks, replacing useless chunks and sewing all the pieces together.”



CONSIDERATIONS FOR TRAINING

- ✗ It's not about TM + MT tasks only
 - + It's about TM-100% + TM-85% + MTopt + MTsubopt
- ✗ Tricks such as “the 5 words rule” might not work
- ✗ Training must consider how to
 - + quickly identify usable sequences
 - + not despair with cumbersome output
 - + delete with no remorse
 - + quickly saw pieces together



WHY, WHEN AND HOW?

× Why?

- + During workshop, even with “low quality” MT output, speed didn’t decrease overall
- + MT seems to activate terminology and provide a model to accept or reject

× When?

- + Domains and text-types

× How?

- + Coupled with TMs
 - × What is the quality threshold where MT hinders productivity?
- + Improvements: customisation possibilities
 - × Terminology
 - × Structures or phrases

THANK YOU!



A DESCRIPTION OF POST-EDITING, FROM TRANSLATION STUDIES TO MACHINE LEARNING

Félix do Carmo and Belinda Maia

Faculdade de Letras da Universidade do Porto

<https://sigarra.up.pt/flup>

Since Holmes' map of the discipline, there has been a space in Translation Studies for research on the translation process. However, when Statistical Machine Translation (SMT) appeared, the description of the translation process in Translation Studies seemed to be overshadowed by the efficiency of algorithms and mathematical techniques to suggest matches and rebuild near functional sentences from big bilingual data. Although it is considered a quality requirement, post-editing is regarded as a very simple process by translation companies and their clients, who require no specification of techniques, but simply usable results.

In this paper, we move away from the concepts of the translation and the post-editing processes as seen by the most current Translation Studies research, in order to integrate the perspective of post-editing in the SMT literature. Inspired by this, we suggest that, in order to distinguish it from Translation and Revision, Post-editing of SMT can be best defined by its technical procedures. Finally, this view of Post-editing leads to the suggestion that the integration of Machine Learning into translation tools should be adjusted in order to support the work of post-editors more efficiently. This paper is based on the experience of one of the authors in managing and doing research in translation and post-editing projects.

A DESCRIPTION OF POST-EDITING, FROM TRANSLATION STUDIES TO MACHINE LEARNING

FÉLIX DO CARMO AND BELINDA MAIA (CLUP)

BARCELONA, OCTOBER 2016

The background is a light blue gradient with several realistic water droplets of various sizes scattered across it. Some droplets are at the top, some at the bottom, and some on the sides. They have highlights and shadows, giving them a 3D appearance.

WHERE IT ALL BEGINS...

TERMS AND CONCEPTS

LEARNING FROM MISTAKES

- Post-edition or Post-editing?
- Learning, teaching...
- Interpreting or Interpretation?

THE STARTING POINT

- James Holmes – Defining the discipline “Translation Studies”
- Process-oriented Descriptive Translation Studies
 - The translation process as a “black-box” mental process
- Applied Translation Studies
 - Teaching
 - Tools:

“A need for scholars in applied translation studies to clarify and define the specific requirements that aids of these kinds should fulfil if they are to meet the needs of practising and prospective translators.”

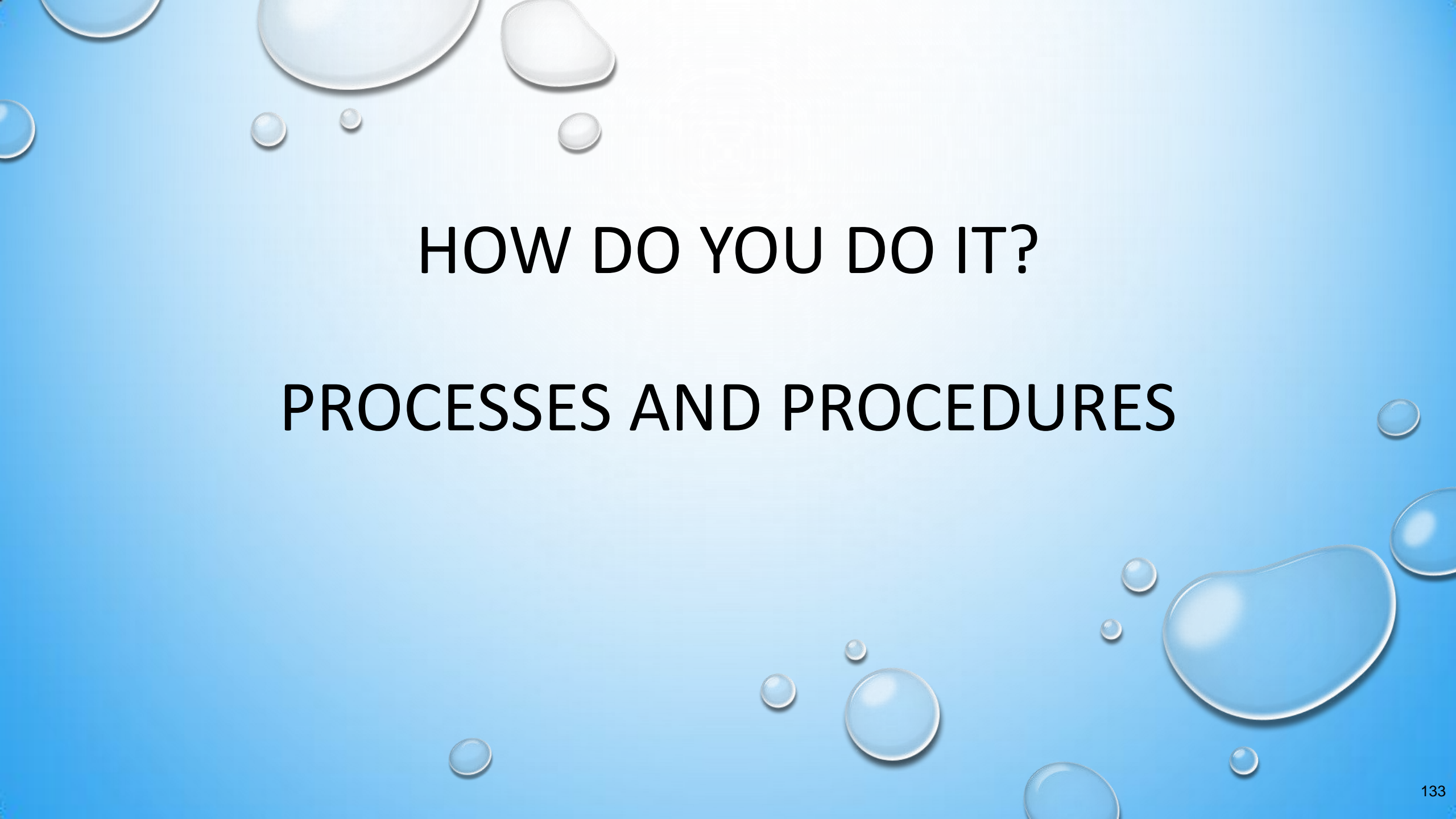
(James Holmes, 1988)

MORE AND LESS THAN TRANSLATION

- **Localization** – more than translation (adaptation)
- **Transcreation** – more than translation (creativity)
- **Post-editing** – less than translation (the translation is done by the machine)

EDITING

- Outside the translation world (cinema, publishing...):
 - Select and prepare material **produced by others** for publication
 - Selecting, deleting, creating meaningful sequences...
 - The “invisible art”
- **INSIDE THE TRANSLATION WORLD**
 - Editing vs Revising (Mossop, 2007 – original texts vs translations)
 - Editing of:
 - Machine-translated text
 - Fuzzy matches
 - Close languages



HOW DO YOU DO IT?

PROCESSES AND PROCEDURES

THE TRANSLATION PROCESS

- INTERNAL VIEW (Descriptive Translation Studies)
 - Cognitive approaches
- EXTERNAL VIEWS (Applied Translation Studies)
 - Teaching and improving how to translate and building better tools
 - Fundamental for professional work, in a demanding industrial context
 - Translation techniques OR Translation solutions (Pym, 2016)
 - “Omissions, additions, changes of location and manipulations of segmentation”
(Gideon Toury, 1995)

TRANSLATING WITH CATS

- Repeated segments – Support at the segment-level
 - Translators validate or edit
- New segments – Support at the word level
 - Concordance and terminology
 - Translators write over the source words
- Fuzzy matches – Support by highlighting differences
 - Translators edit
 - And re-edit (internal fuzzy matches – similar segments edited in the source text)

POST-EDITING WITH CATS

- Repeated segments – CAT tool support
 - Translators validate or edit
- New segments – Machine-translation support
 - Translators post-edit machine-translated segments
- Fuzzy matches – CAT tool support
 - Translators edit and re-edit

BUT

- Post-editing of **New segments** is made with the same CAT word-level support
- Although most of the editing effort is in **Fuzzy matches**, there is no help from MT

IMPROVED SUPPORT IN CATS

- Fuzzy match composition
 - Or “fuzzy match repair”, or “advanced leveraging”
 - Translators edit translation suggestions built from sub-segment solutions retrieved from bilingual data (terminology or alignment data)
 - The challenges of supporting fuzzy matching with MT (Esplà-Gomis et al, 2015)
- Predictive writing
 - Autocomplete, type-ahead procedures
 - Interactive – suggestions appear as you type the first characters of words

The background is a light blue gradient with several realistic water droplets of various sizes scattered across it. Some droplets are at the top, some at the bottom, and some on the sides. They have highlights and shadows, giving them a 3D appearance.

CAN WE DO IT BETTER?

**IMPROVING POST-EDITING WITH
MACHINE LEARNING**

EDITS IN MT

- **EDIT DISTANCE**

- Levenshtein (1966): deletion, insertion and substitution
- Damerau (1964): deletion, insertion, substitution and transposition

- **TER – Translation Edit Rate**

- Snover (2006) – a metric to measure the quality of MT results
- Minimum number of edits (Delete; Insert; Move/shift; Replace) divided by the number of words

- **Machine Translation Quality Estimation**

- Specia et al (2010) – to identify whether the MT system will produce a good-enough hypothesis to be worth post-editing
- Identify the features that make this estimation possible:
 - Complexity of the source
 - Fluency of the target
 - Adequacy of the translation
 - Confidence in the MT system
- **Estimate editing effort, in terms of 4 editing tasks**

LEARNING FROM POST-EDITING

- First generations – batch learning
 - New data: new training of the models
- Online learning
 - Especially active since MateCAT and CasMaCAT projects
 - “Interactive machine translation” or “Interactive translation prediction”
 - The system builds a hypothesis, the user types, the system presents contextual suggestions
 - Aim: to improve the TM system and thus to reduce translation effort
 - Editing support is based on the predictive writing paradigm

Based in (among others) Ortiz-Martinez (2016)

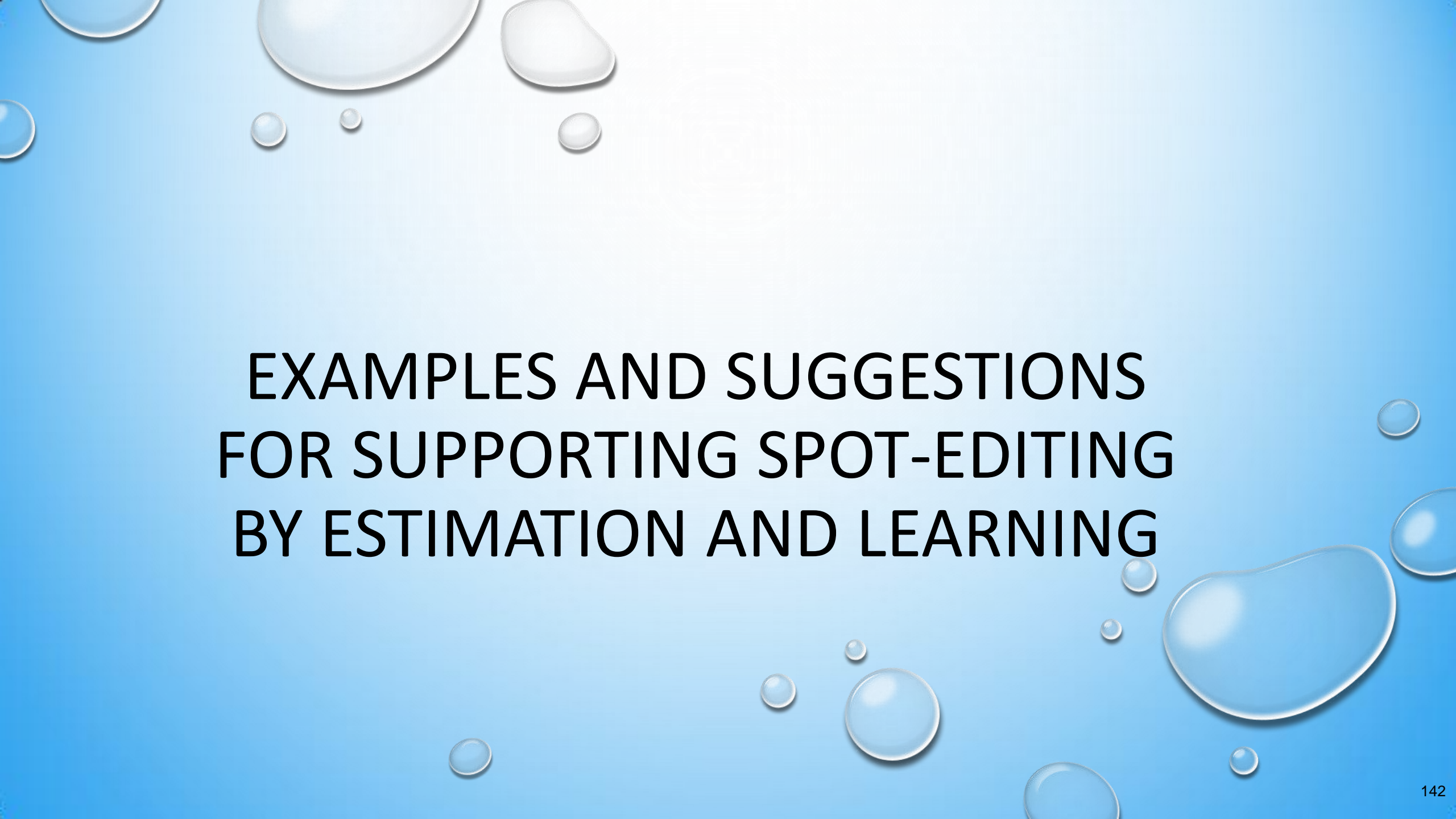
INTERACTIVE SUPPORT TO POST-EDITING

- Current
 - Focused on improving MT
 - Based on predictive writing

BUT isn't that an extension of translation support?

In Post-editing, typing is more scattered - not as linear as in translation (Carl et al, 2015)

- **Our project**
 - Focus on helping translators do what they do when they post-edit, based on the 4 editing tasks
 - Based on work done on Quality Estimation to estimate editing
 - Sustained by a “source text model” (an analysis of the repeated phrases in the source text)
(Bertoldi et al, 2014)
 - Based in Machine Learning to improve models with translators' actions

The background is a light blue gradient with several realistic water droplets of various sizes scattered across it. Some droplets are at the top, some at the bottom, and some on the right side. The text is centered in the middle of the slide.

EXAMPLES AND SUGGESTIONS FOR SUPPORTING SPOT-EDITING BY ESTIMATION AND LEARNING

INTERACTIVE DELETION

SOURCE	MT SUGGESTION	POST-EDITED
Acquire - to obtain possession of something	Adquirir - para obter a posse de algo	Adquirir - obter a posse de algo
Align - to place something in an orderly position in relation to something else	Alinhar - para colocar algo em uma posição ordenada em relação a outra coisa	Alinhar - colocar algo em uma posição ordenada em relação a outra coisa
Allocate - to divide something between different people or projects	Alocar - para dividir algo entre diferentes pessoas ou projetos	Alocar - dividir algo entre diferentes pessoas ou projetos

- Estimating deletion:
 - Word/units in chosen hypothesis BUT which have a low occurrence in that context in the target Language model (besides repeated words, untranslated words...)

INTERACTIVE INSERTION

SOURCE	MT SUGGESTION	POST-EDITED
User Name/ID	Nome de utilizador	Nome/ ID de utilizador
Patient Name/ID	Nome do paciente	Nome/ ID do paciente
Item Name/ID	Nome do item	Nome/ ID do item

- **Estimating insertion**
 - Words that are frequent in that context in the target Language model, but which are missing in the translation hypothesis (articles, titles (Sr., Dr.), etc.)

INTERACTIVE MOVEMENT

SOURCE	MT SUGGESTION	POST-EDITED
VEC 1 controller pin 7 (BK) wire	Controlador VEC 1 fio do pino 7 (BK)	Fio do pino 7 (BK) do Controlador VEC 1
VEC 1 controller + (RD) wire	1 Controlador VEC + (RD)	Fio + (RD) do Controlador VEC 1
VEC 1 controller – (BL) wire	VEC 1 controlador - (BL)	Fio - (BL) do Controlador VEC 1

Learning sub-segment translation units

- “Matching” in “pre-postediting” (Marie and Max, 2015)
- Validated when translators **Move** translation units
- Add to translation model with a higher (101%)confidence score

INTERACTIVE REPLACEMENT

SOURCE	MT SUGGESTION	POST-EDITED
Users must be set up and maintained at the console.	Os utilizadores têm de estar configurado e mantido na consola.	Os utilizadores têm de estar configurados e mantidos na consola.
Assess - to examine something in order to judge or evaluate it	Avaliar - examinar algo para juiz ou avaliar	Avaliar - examinar algo para ajuizar ou avaliar
Act - to do something to change a situation	Ato - fazer algo para mudar uma situação	Atuar - fazer algo para mudar uma situação

- Interactive Replacement
 - Keep and present lower rated alternative hypotheses in translation tables
 - Or inflected forms of selected units (based on inflexion rules)

FINAL REMARKS

TRANSLATION, REVISION AND POST-EDITING

- These are all tasks in a translation workflow, that may coexist in the same project
- They all involve some degree of editing
- In future, both **Fuzzy matches** and **New segments** will be presented as a combination of fuzzy match composition and Machine Translation
- Post-editing may be seen as being restricted to the 4 editing tasks, concepts that come a long way back from both Translation Studies and Machine Translation literature.

CONDITIONS FOR POST-EDITING

- The source text should have an adequate degree of Complexity (internal phrase repetition)
- The MT system should have an adequate degree of Confidence
- The translation tools should give the translator support for the four editing tasks
- The translator should be able to do no more than a reasonable number of edits
 - Suggestion: 25% edit distance (as for fuzzy matches in CATs)
 - Translated words – above 25% editing
 - Edited words – below 25% editing

POST-EDITING: AN AUGMENTED DEFINITION

A process applied to machine-translated text, in order to improve its quality, composed of 4 editing tasks (delete, insert, move and replace), within a specific threshold.

As for **Translation**, it is so much more than that...

THANK YOU.

REFERENCES

- BERTOLDI, Nicola, Patrick Simianer, Mauro Cettolo, Katharina Wäsche, Marcello Federico, and Stefan Riezler. 2014. "Online adaptation to post-edits for phrase-based statistical machine translation." in *Machine Translation* 28, 3-4 (December 2014), 309-339.
- GARCIA, Ignacio. 2012. "A brief history of postediting and of research on postediting." in *Revista Anglo Saxonica*, vol 3, no 3. pp 291 - 310.
- GREEN, Spence, Jason Chuang, Jeffrey Heer, and Christopher D. Manning. 2014. "Predictive Translation Memory: A Mixed-Initiative System for Human Language Translation". In *UIST*.
- HOLMES, James S. (1988). *Translated!: Papers on Literary Translation and Translation Studies*. Amsterdam: Rodopi
- MOSSOP, Brian. 2007. *Revising and editing for translators* – 2nd Edition. Manchester: St. Jerome.
- SNOVER, Matthew, Dorra, B., Schwartz, R., Micciulla, L., and Makhoul, J. 2006. "A Study of Translation Edit Rate with Targeted Human Annotation". in *7th Conference of the Association for Machine Translation in the Americas*, pages 223–231, Cambridge, Massachusetts.
- SPECIA, Lucia, Dhvaj Raj, and Marco Turchi. 2010. "Machine translation evaluation versus quality estimation". in *Machine Translation* 24, 1 (March 2010), 39-50.
- TOURY, Gideon (1995). *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins
- PYM, Anthony. 2016. *A typology of translation solution types*. Web: <https://www.academia.edu/s/f91e5c89bb/a-typology-of-translation-solution-types>
- ESPLÀ-GOMIS, Miguel, Felipe Sánchez-Martínez, and Mikel L. Forcada. 2015. "Using machine translation to provide target-language edit hints in computer aided translation based on translation memories". in *Journal of Artificial Intelligence Research*, vol. 53, 1 (May 2015), 169-222.
- LEVENSHTAIN, Vladimir I. 1966. "Binary codes capable of correcting deletions, insertions, and reversals". In *Soviet Physics Doklady*, 10:707–710.
- DAMERAU, Fred J. (March 1964), "A technique for computer detection and correction of spelling errors", in *Communications of the ACM*, ACM, **7** (3): 171–176.
- ORTIZ-MARTÍNEZ, Daniel. 2016. "Online Learning for Statistical Machine Translation". in *Computational Linguistics* 42(2):1-41.
- CARL, Michael. Isabel Lacruz, Masaru Yamada, Akiko Aizawa: "Comparing spoken and written translation with post-editing in the ENJA15 English → Japanese Translation Corpus." in *The 22nd Annual Meeting of the Association for Natural Language Processing (NLP2016)*, Tohoku University, Japan, 2016.03.
- MARIE, Benjamin and Max, A. 2015. "Touch-Based Pre-Post-Editing of Machine Translation Output". In *EMNLP 2015*, Lisbon, Portugal.

MACHINE TRANSLATION: A SECRET LOVER TURNED INTO THE LIFE PARTNER

Diego Bartolomé

tauyou language technology

www.tauyou.com

No matter if we like it or not, Machine Translation (MT) is here to stay. Certainly, there are several fields where MT (nor CAT tools) will never be used because the process of creating a translated text requires a significant creativity. However, it is currently used in many technical fields, and also for verticals such as e-commerce where in many cases a machine translated sentence plus a light post-editing is good enough for the end client. However, don't worry, top quality translations will never disappear! But some other types of translation will be replaced by Big Data processing and natural language processing ...

In this presentation, we will show why MT might not be well-suited for certain fields, but also the good translations produced by MT in fields such as e-commerce or hospitality with Big and Small Data. Then, the optimum process to deal with post-editing jobs from the freelance translator point of view will be described, discussing key aspects such as the discount or how to deal with complaints. Finally, we will cover how tools in the market deal with machine translation post-editing, and some products specifically designed for the task.

In the end, we will show that if MT helps freelance translators increase their profits or have more free time, and also opens new business opportunities, the optimum option we all have is to embrace change and love MT.

Machine Translation: The Secret Lover turned Life Partner

Diego Bartolome
@diegobartolome

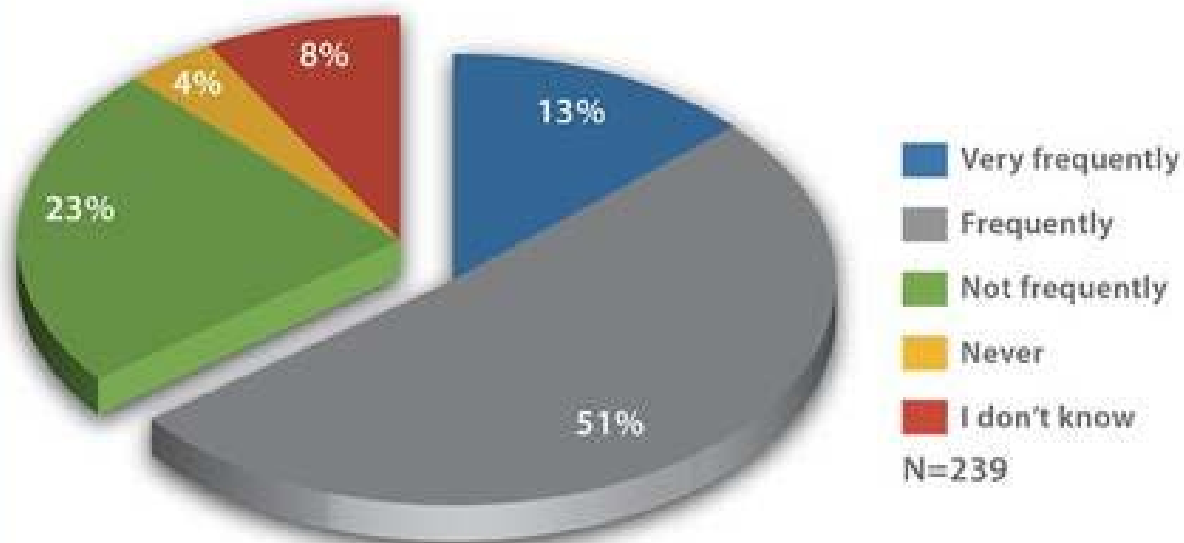
Google

bing™

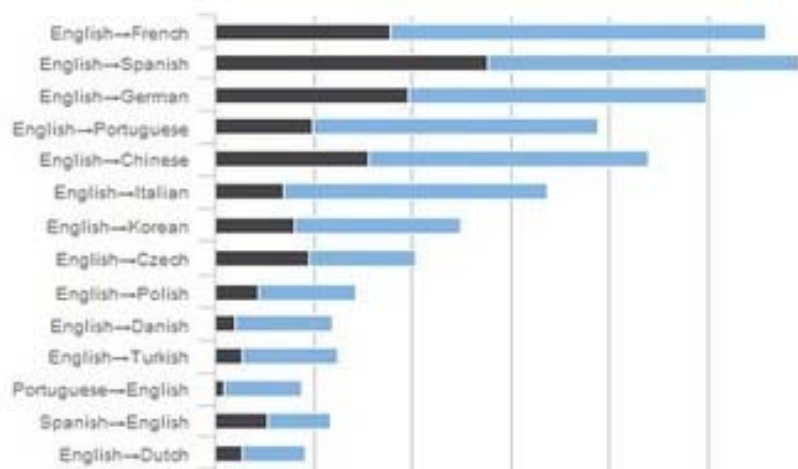
Yandex

Baidu 百度

How often do you think your colleagues use free machine translation?



Source: Common Sense Advisory, Inc.

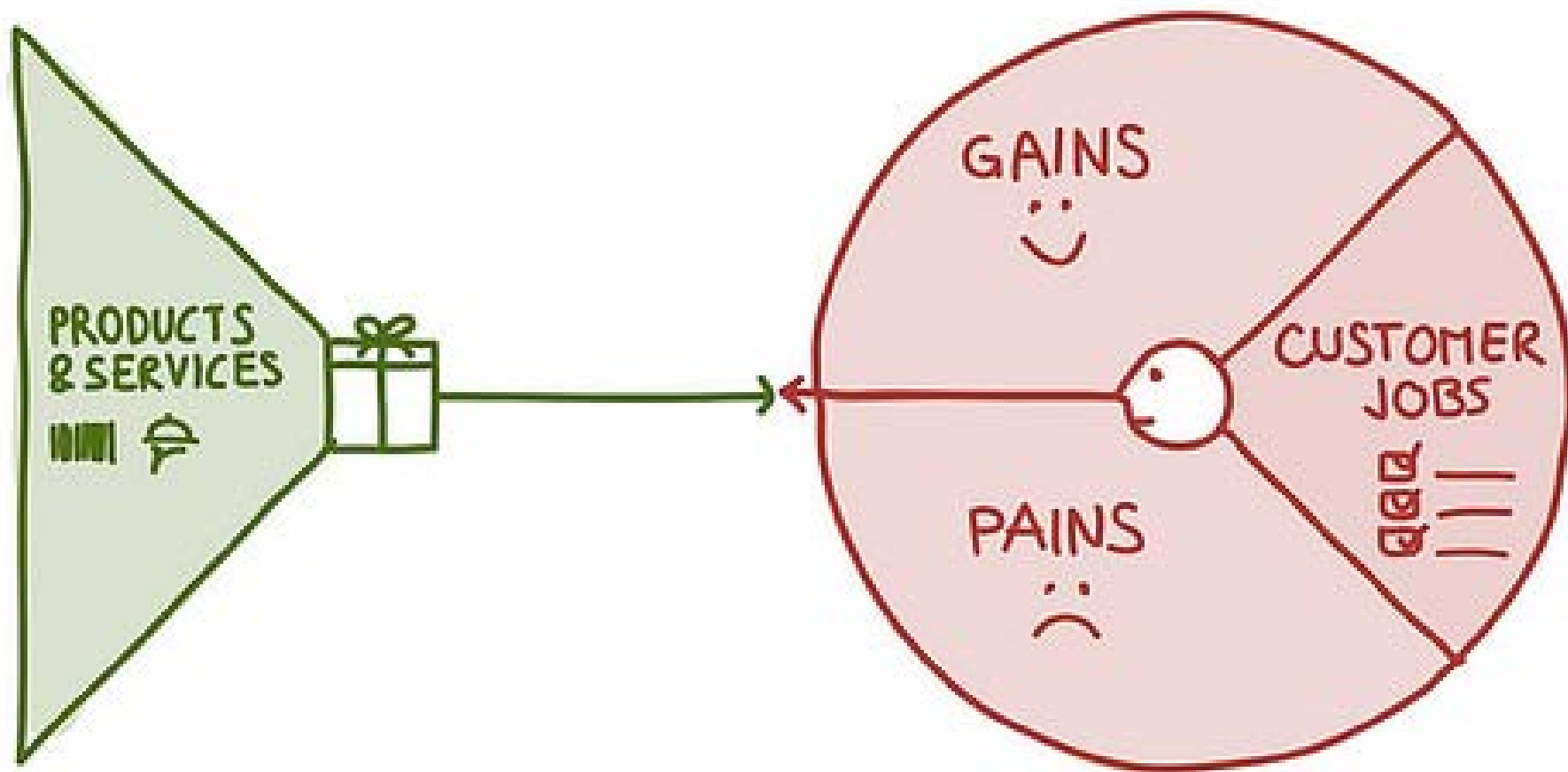


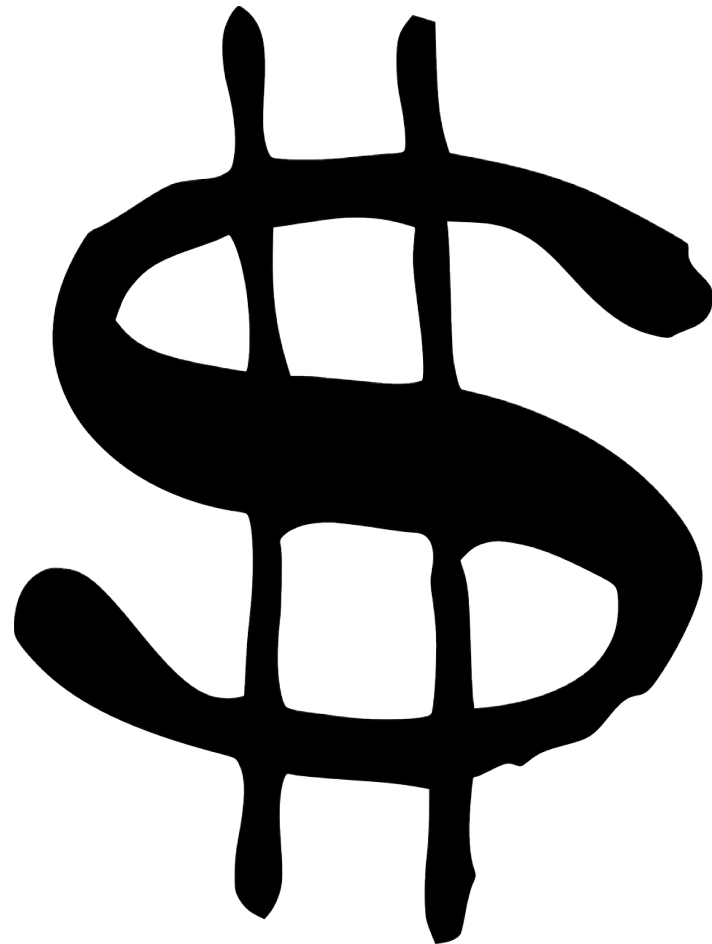


50%









LIKE

1. Decision on strategy of engine
2. Training of first version
3. Objective quality evaluation
4. Implementation of post-editing rules
5. Estimation of translation speed
- 6. Change Translation into MT post-editing**
7. Implementation of linguistic feedback
8. Update of the engine

The Discount

cloudl.memsource.com/web/analyse/show/10901

Download LOG Calculate Discount

Total files analyzed: 1

de_de → it_it

File: Agreement.docx

	Segments			Pages			Words			Characters			%		
	TM	MT	All	TM	MT	All	TM	MT	All	TM	MT	All	TM	MT	All
All	0	9	9	0	0.5	0.5	0	110	110	0	917	917	0	100	100
Translations	-	-	0	-	-	0	-	0	0	-	-	0	-	-	0
100%	0	9	9	0	0.5	0.5	0	110	110	0	917	917	0	100	100
99%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
85%–94%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

When no post-editing is done, segments come up as 100% MT matches.

With all 110 segments in the 100% range, either the MT is perfect or the post-editor did not do the job properly.

Statistics and reports

When you talk about statistics, in most cases you can estimate your cost or time requirements. To set different options enables you to estimate the users of competitive products.

CAT tools use statistical language processing to find similar segments are also similar in meaning, for data in the statistical analysis in different CAT

memoQ examines the segments, looks up the data of documents and the report options.



Post-Edit Compare

By Patrick Hartnett

Version See Codificare.net

Post-Edit Compare is a tool designed to report translation modifications during the post-edit phases of a translation workflow.

The tool works by comparing two versions of the same SDLXLIFF file (before and after changes were applied); the files with changes are highlighted and all modifications are included in a comparison report.

The comparison report is formatted in a way that simplifies the understanding of what changes were applied from one version of the file to the other, along with a full break down of the modifications and related cost analysis.

Once installed, you will now find Post-Edit Compare as a desktop shortcut. In addition, upon restarting SDL Trados Studio, you will now find in the navigation menu on the left of the interface, a new option called 'Post-Edit Versions'.

Error Reports

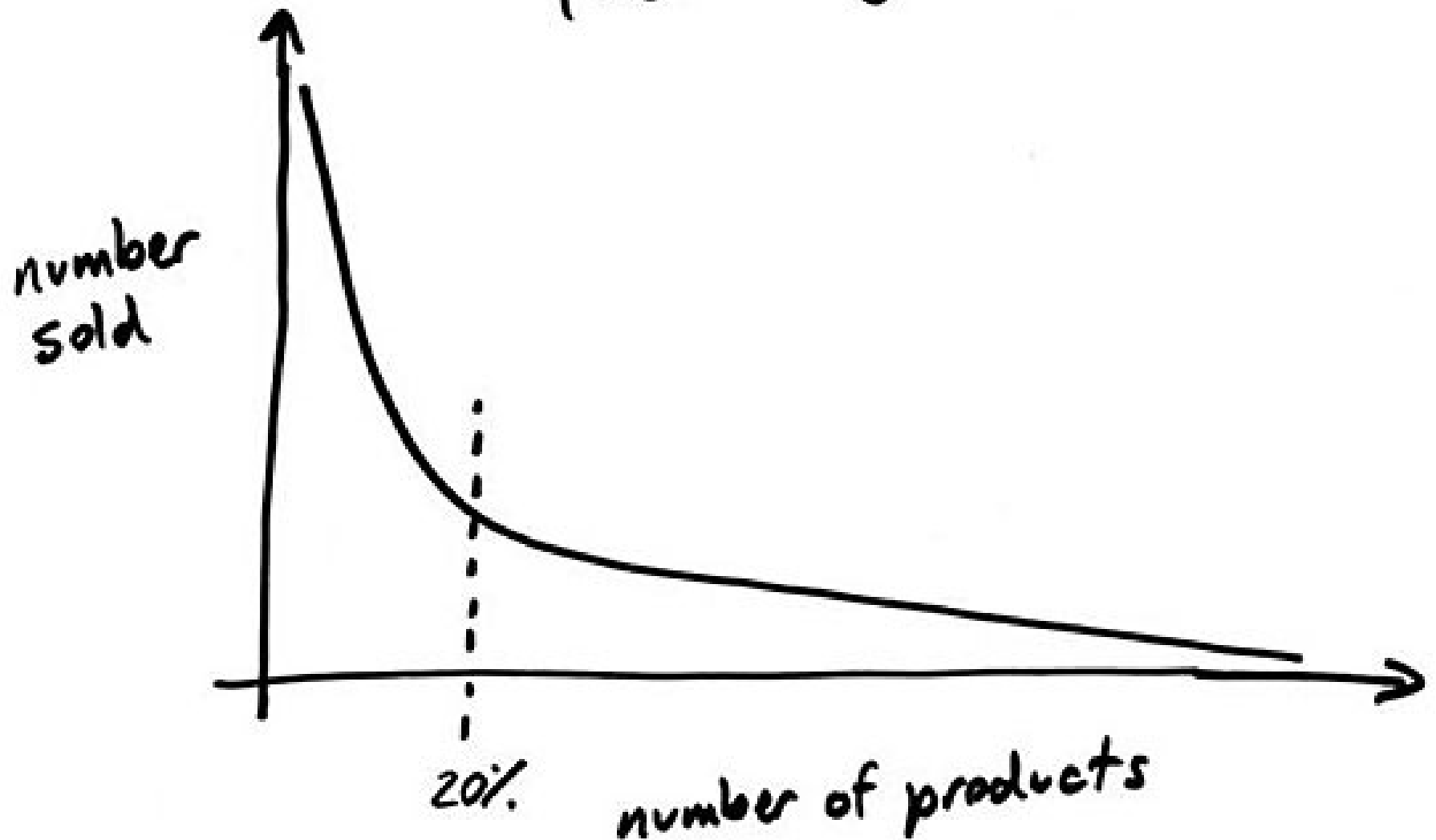
Subjective feedback → most common errors

Detailed report of repetitive mistakes

Time-wasting errors

New Business Opportunities

The "Long Tail"





METHODOLOGY OF MT POST-EDITORS TRAINING. PRACTICAL EXPERIENCE WITH PEMT AS A PART OF TRANSLATION STUDIES AT CONSTANTINE THE PHILOSOPHER UNIVERSITY IN NITRA.

Jakub Absolon

Constantine the Philosopher University in Nitra

www.Asap-translation.com

In my presentation I would like to explain why we understand there is a need for specific methodology for PEMT training, based on experience both with professional translators and students.

Current students are more than familiar with statistical based machine translation and therefore it should be reflected in methodology we use. They are “natural-born Googlers” and that is why they should understand how MT can help them but also how it could threaten them. Our main interest is to find out which are the most important competencies for translators using PEMT as a natural technique of translation process and what are the differences (if any) in comparison to a traditional translator. The second goal is to find out which of the competencies of PE are trainable, to what extend and what is the best methodology for training.

Practical experience with PEMT as a part
of Translation Studies at Constantine the
Philosopher University in Nitra

Jakub Absolon

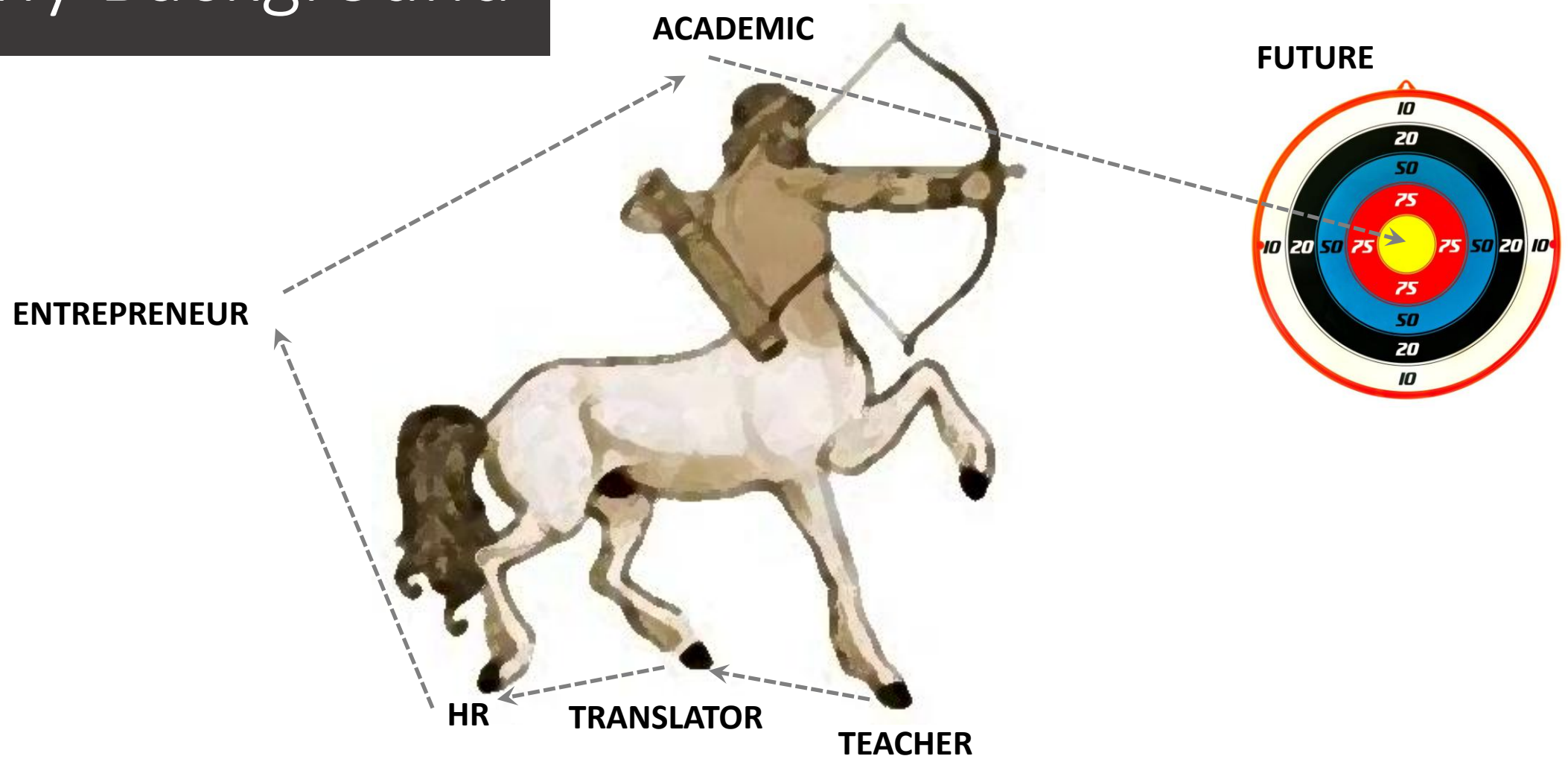
Methodology of MT Post-Editors Training

Heavy
WordLoad

Translator
Post-Editor

MT Engine

My Background



PEMT as a Part of Translation Training at UKF

Courses with content connected with PEMT

Goals:

- Make principals of MT known
- Basics of post-editing of machine translation
- Error typology, the most frequent EN-SK errors made by MT engines
- PEMT practice as part of internship program

Course/Module	Hours week/semester	Course/Module content
Computer Aided Translation	2 per week / 26 per semester	Presentation of the current technical reality and its interaction with translation techniques, schematization of the translation workflow model; development of skills for modern computer tools supporting the translation act. The course also covers the work with translation memory software and terminology management systems. Outcome: methodological and technical background for utilization of information systems supporting the translation act.
Machine Translation	2 per week / 26 per semester	The goal of this course is to get students acquainted with the newest trends in the field of machine translation. Both quantitative and synthetic models of MT are taken into account and thoroughly dealt with. Extensive exercises in MT post-editing are also part of the course.
Computational Linguistics	3 per week / 39 per semester	The central goal of this course is to familiarize students with core techniques and applications of computational linguistics. Students will gain an appreciation for methodology for empirical linguistic analysis and natural language processing involving the use of text corpus, language modelling etc.



Why PEMT Research

It's trendy 😊.

It's present and future of technical translation and we would like to know who are the best PEs.

Who are the best post-editors?

- Good translators = good post-editors
- Professionals from other fields are better post-editors than translators
- „The younger the better“ - novice translators and translation studies students
- Amateur crowd – it doesn't matter who, „just do it“ principle
- WE



People Still Matter in PEMT

Philipp Koehn
(Johns Hopkins University),
MT Marathon 2016 Prague:

„There is higher variability
between translators than
between MT engines“

=

Selection and training of
post-editors matters!!!



Translator Variability

27



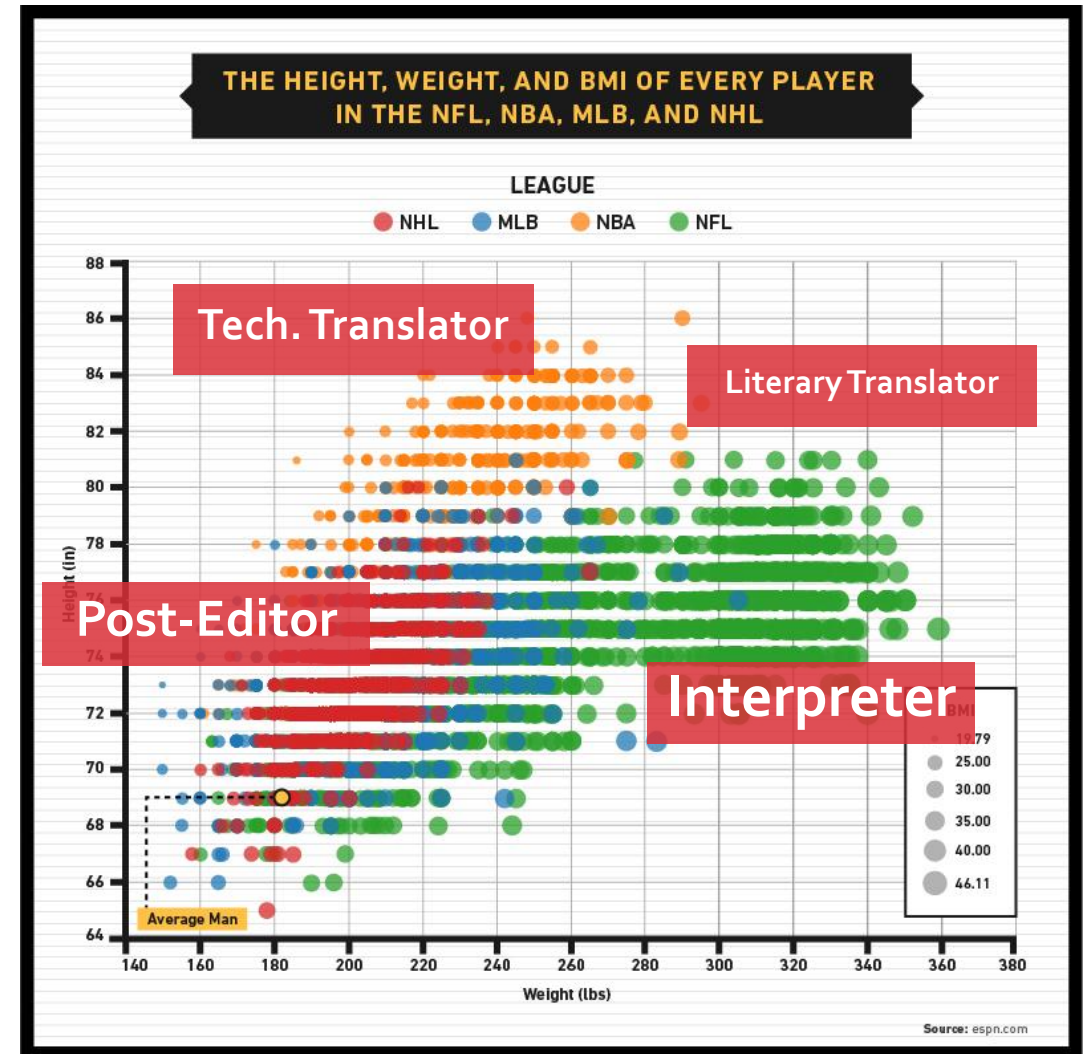
	HTER	Edit Rate	PE speed (spw)	MQM Score	Fail	Pass
TR1	44.79	2.29	4.57	98.65	10	124
TR2	42.76	3.33	4.14	97.13	23	102
TR3	34.18	2.05	3.25	96.50	26	106
TR4	49.90	3.52	2.98	98.10	17	120
TR5	54.28	4.72	4.68	97.45	17	119
TR6	37.14	2.78	2.86	97.43	24	113
TR7	39.18	2.23	6.36	97.92	18	112
TR8	50.77	7.63	6.29	97.20	19	117
TR9	39.21	2.81	5.45	96.48	22	113

- Higher variability between translators than between MT systems



Challenges & Questions

- **Identification of differences** in job profiles of translator vs. post-editor
- **Trainability** of PE skills and competences
- Recommended **methodology** for PEMT (e.g. split techniques)
- Necessary **workload** for PEMT training





Source: <http://www.simplyshredded.com>

Our Principals

Competency-based approach

- “Competency is an underlying characteristic of an individual predicting behaviour and performance.”

Split training

- Lesson learned (How Arnold turned his biggest weakness - his calves into a showcase muscle)
- Splitting is the way go. Full body training, provided that the intensity is high and the routine is good, can produce some amazing results, but splitting lets you get more from less: more results in a shorter period of time spent in the gym.

Source: <http://www.bodybuilding.com>



Competency-Based

- We focus on core competences
- EMT competency model:
 - Language competence
 - Intercultural competence
 - Information mining competence
 - Technological competence
 - Thematic competence
 - Translator service provision competence



Source: EMT



Split Training

- **Faster results**
- **Assess Your Weaknesses**
- Split training = exercises for specific competencies (e.g. language, technical, thematic)
 - Exercises for specific error types (language competency):
 - Morphology / Grammar (discord in case, number, part of speech,...)
 - Typography (capital letters, punctuation,...)
 - Syntax (word order)
 - Meaning / Semantics (wrong translation, missing translation)
 - Terminology (usage of general vocabulary instead of domain specific)
 - Facts (numbers, dates,...)



Research Project

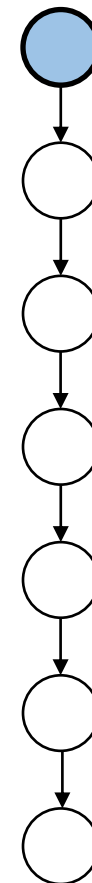
- A. **Pilot project** - survey on attitudes and practice of PEMT usage by students of Translation Studies.
- B. **Observation and analysis of MT post-editor's work** and its comparison with technical translator's (HT) work.
- C. **Post-editor job-profile** creation.
- D. **Designing a training plan** for the development of appropriate post-editing skills consisting of theoretical and practical part.
- E. **Setting up a battery of tests** to objectively measure performance of the post-editor.
- F. **Comparing performance** of HT vs. PEMT search and cross-correlation. (pilot project + professional translators test)
- G. **Verify** the extent to which it is possible to develop MT PE's performance by training plan.



Who were our Respondents?



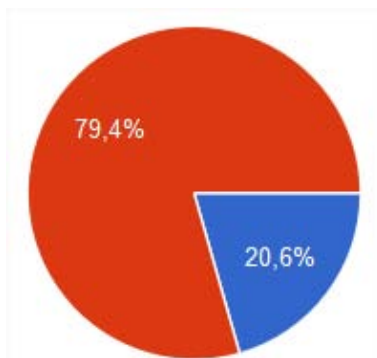
Pilot Project



Pilot Project

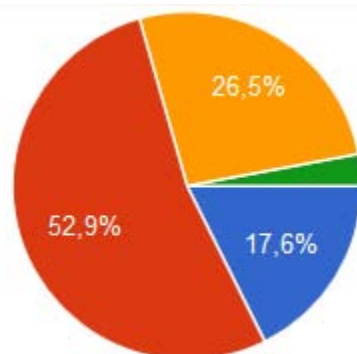
Pilot Project

Translation experience



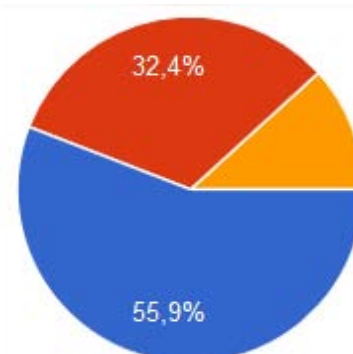
No, I don't translate
(only at school) - 79.4%
I translate at least 5
pages per week - 20.6 %

PEMT awareness
and experience



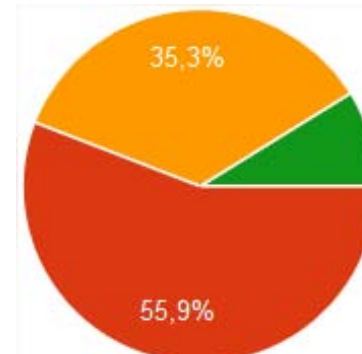
I've never heard of PEMT
17.6%
Personally no, but I've
heard about PEMT
52.9%
Yes, but I have no
practical experience
26.5%
I have practical
experience with PEMT
2.9%

MT usage

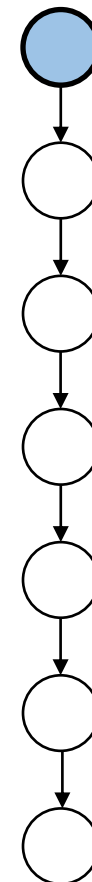


I only use MT for
unknown languages
11.8%
I translate unknown
words 55.9%
I translate a whole
sentence and then edit
32.4%
I don't use it at all 0%

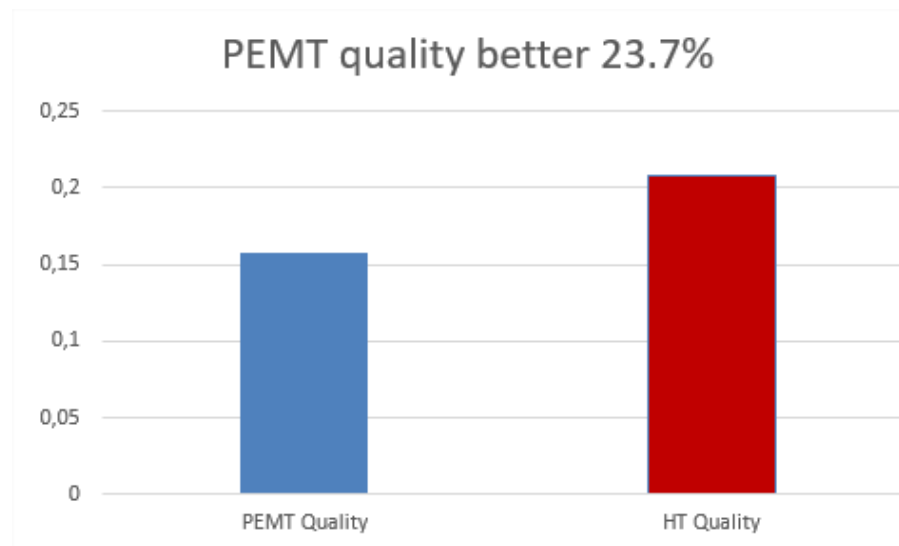
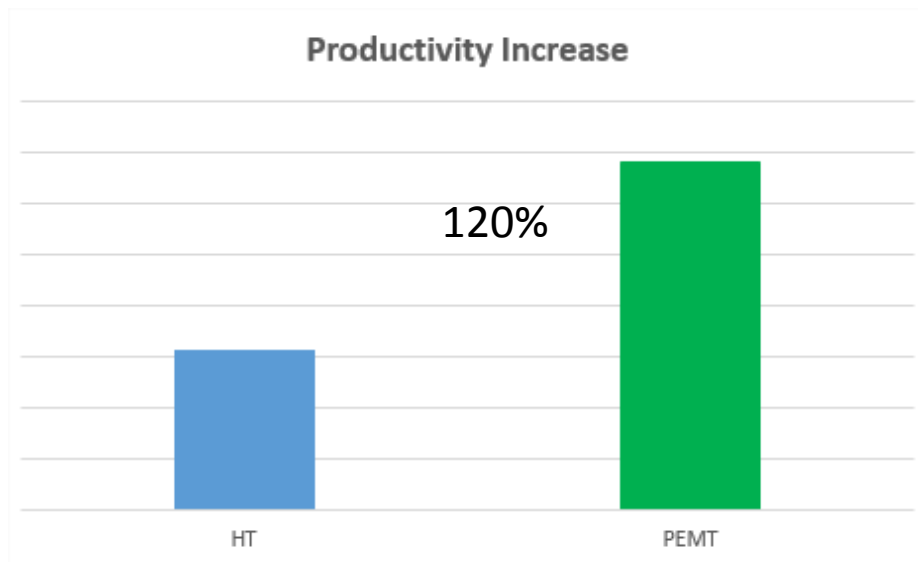
MT quality awareness



It provides high quality
translation 0%
MT quality varies but is a
big help 55.9%
MT helps but it mustn't
be automatically copied
to target 35.3%
MT slows down the
translation process 8.8%
MT is absolutely
unacceptable 0%

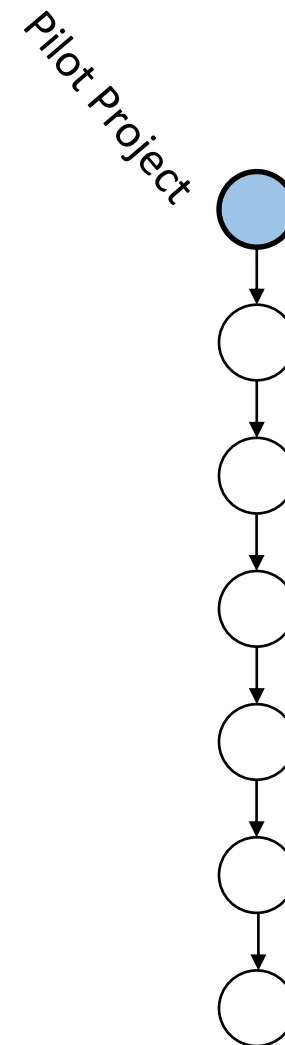


Pilot Project



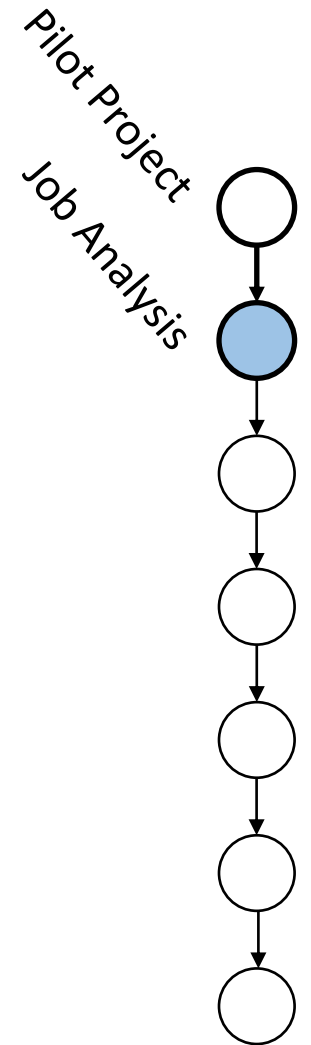
Potential reason:

- Using MT is second-nature for young people
- Well-selected text for MT
- Students weren't experienced translators

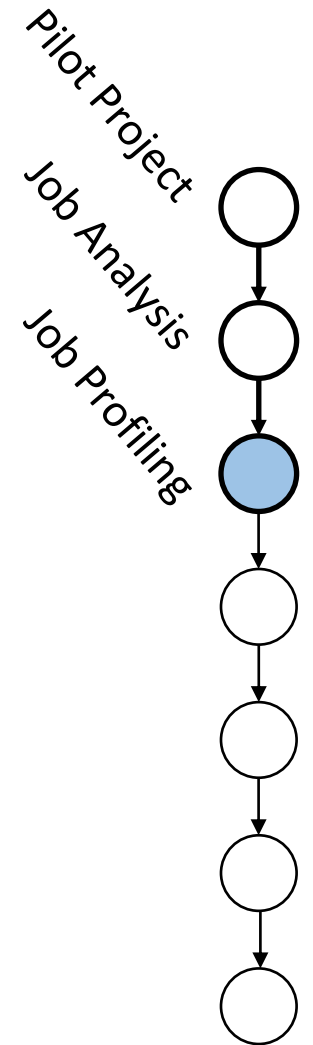
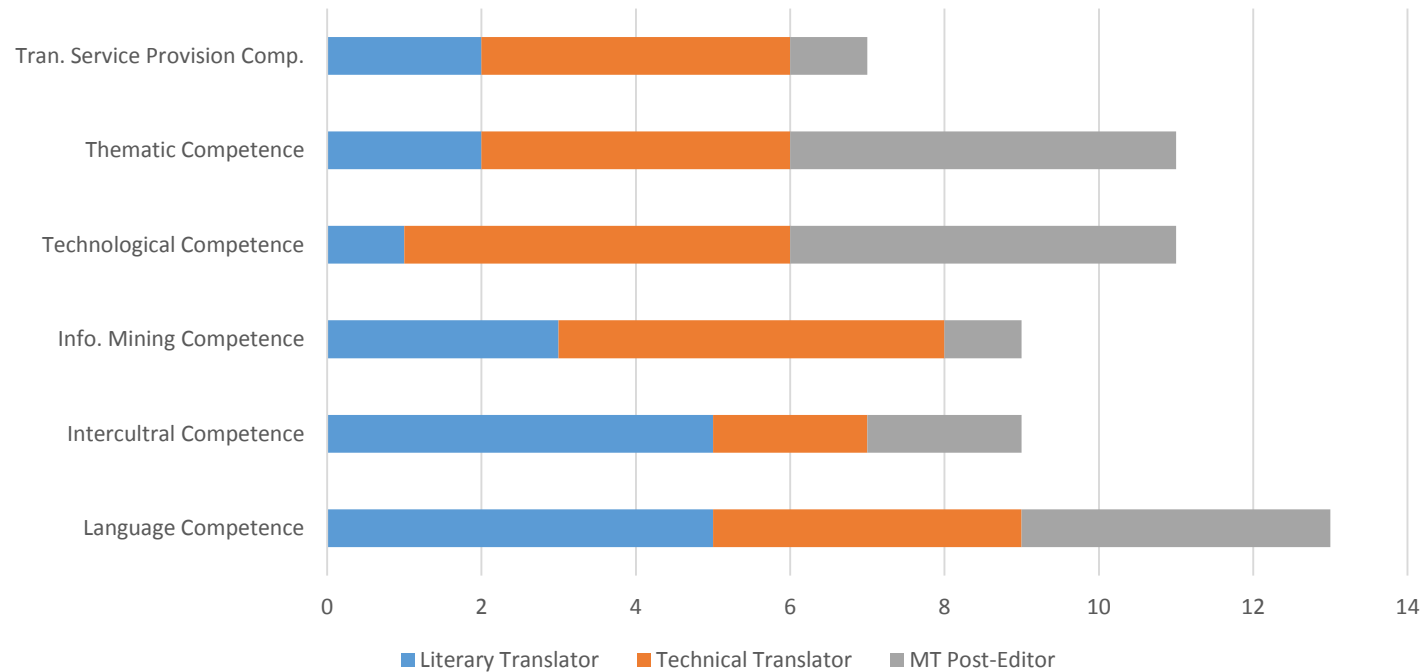


Observation and Analysis of PEMT work

TRANSLATOR	POST-EDITOR
<ul style="list-style-type: none"> • Reads the text in the source language for the purpose of a thorough understanding • Creates translation in the target language • Searches for terminology in the dictionaries and internet • Works with specialized vocabulary (search) • Considers the target audience • Complies with the formal language rules 	<ul style="list-style-type: none"> • Reads machine translation • Fixes the translated text on a linguistic level (word order, grammar,...) • Controls usage of supplied terminology • Compares the original text to translation • Creates new translation (in case of gross errors in MT) • Uses specialized vocabulary



PEMT Job Profiling



Designing a Training Plan

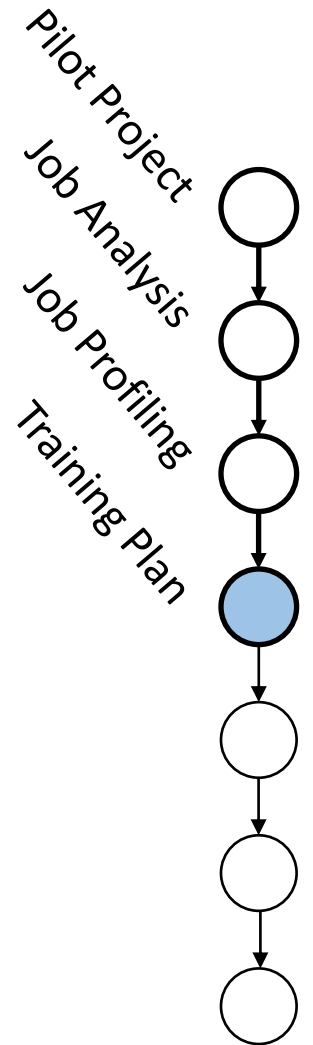
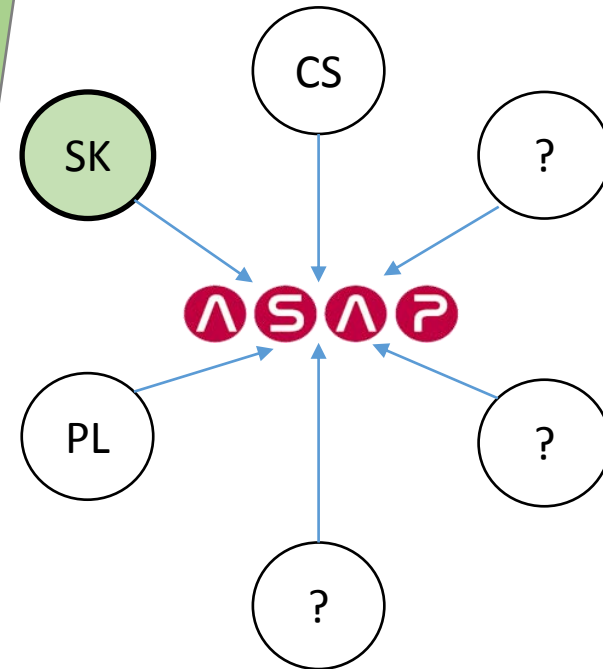
for the development of appropriate post-editing skills consisting of theoretical and practical part



- Research
- PEMT training for students



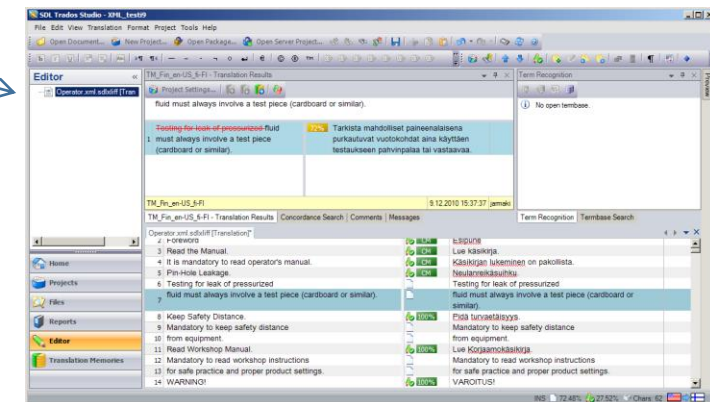
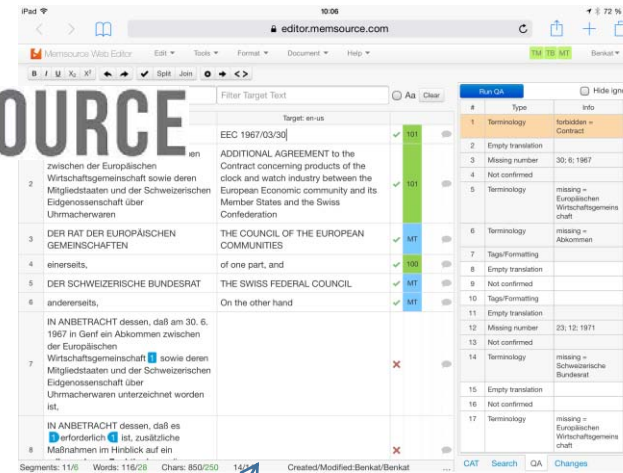
- Test production
- Commercial testing & Training



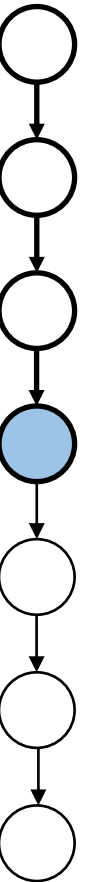
Designing a Training Plan

Exercises:

1. PEMT overview
2. PEMT (Language Competence, Technical Competence)
 - Decision making (Lang. + Tech. + Personality)
 - Error specific PEMT (Language Competence)
 - Error identification (Language Competence)
 - PEMT with prompt (Technical Competence)
 - Free PEMT (Language Comp., Tech. Comp.)
3. Product specific training (Thematic Competence)
 - Terminology training
 - Reading materials for topic comprehension (source and target language)



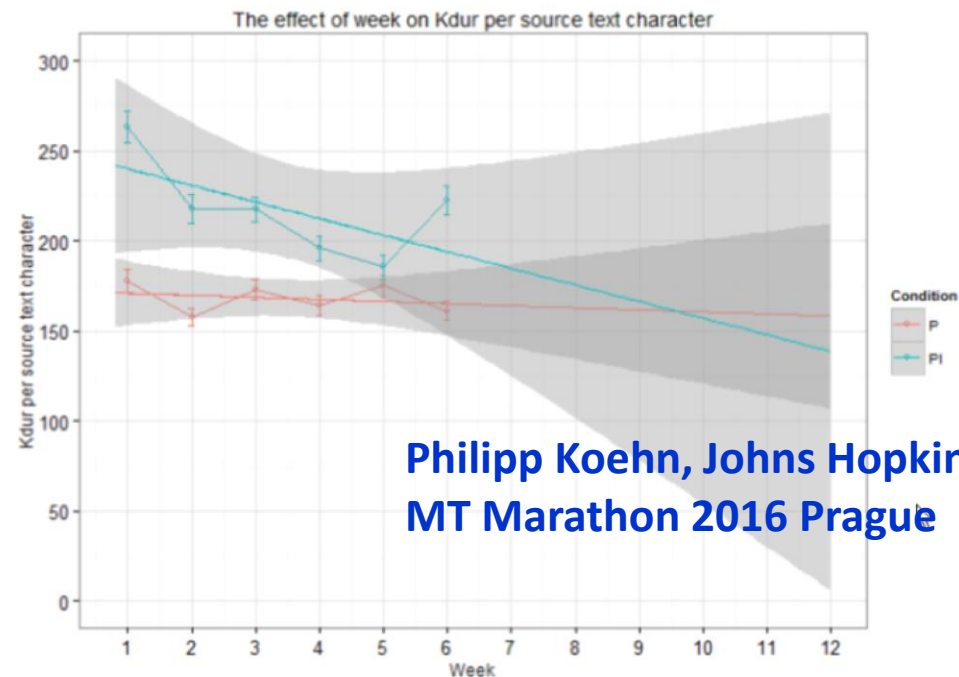
Pilot Project
Job Analysis
Job Profiling
Training Plan



Longitudinal Research

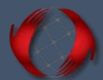
- 10 weeks training (Koehn)
- 10 weeks = 100-200 standard pages
- Our research: 50 standard pages

Learning Effects: Professional Translators 94



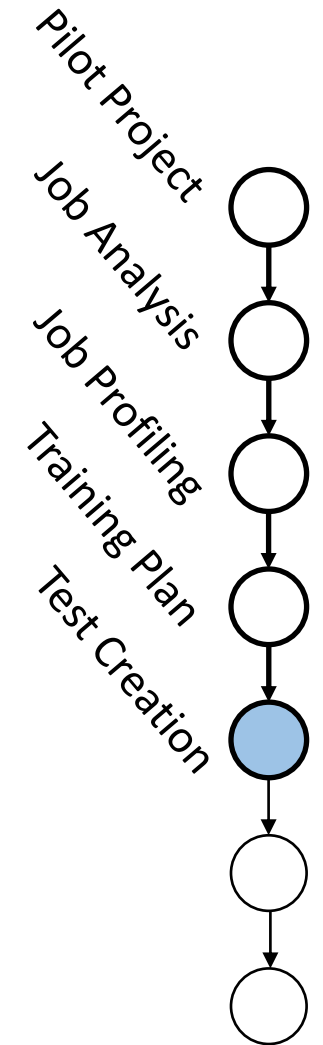
**Philipp Koehn, Johns Hopkins University,
MT Marathon 2016 Prague**

CASMACAT longitudinal study
activity projection as reflected in Kdur taking into account six week
(Kdur = user activity excluding pauses > 5 seconds)



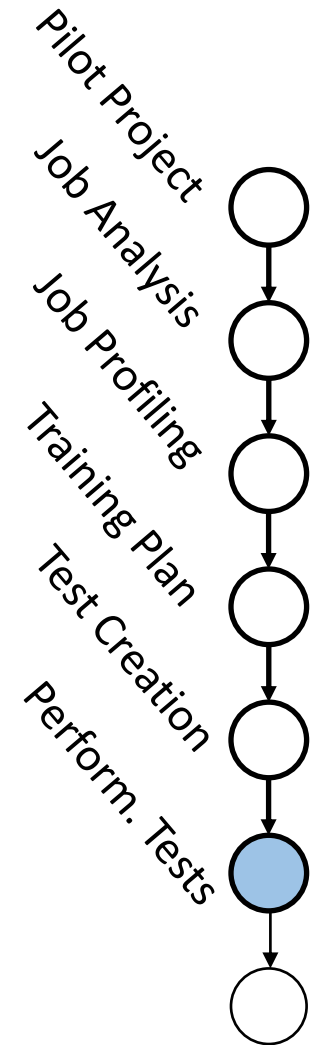
PEMT Performance / Selection Test Creation

TEST	FOCUSED AT COMPETENCE	DESCRIPTION
Target text review	Language	Intentional grammatical errors in the text, which must be identified. Criterion for passing the test is 90% accuracy.
Find differences in text strings	Technical and Personality	The goal is to find the differences in the parallel text for the shortest time. The first part of the text is general text(addresses) and the second part of the test are random strings of characters. Evaluate the accuracy of the test. Differences that identify missed are considered errors. Criterion for passing the test is 90% accuracy.
Terminology test	Thematic	Multiple choice test. It assesses the percentage of accuracy. Criterion for passing the test is 90% accuracy.
Test comprehension of the source text (domain specific)	Thematic and Language	Multiple choice test focused on understanding the text
PEMT speed test	Technical	We measure speed and accuracy of PEMT. Accuracy is evaluated automatically by analyzing the final text of a pretreated translation memory, which contains the target text. Speed is measured based on number of translated words. Criterion for passing the test is 90% accuracy and speed PEMT least 500 words per hour. Prepared text contains sentences that require editing a maximum of 25%.



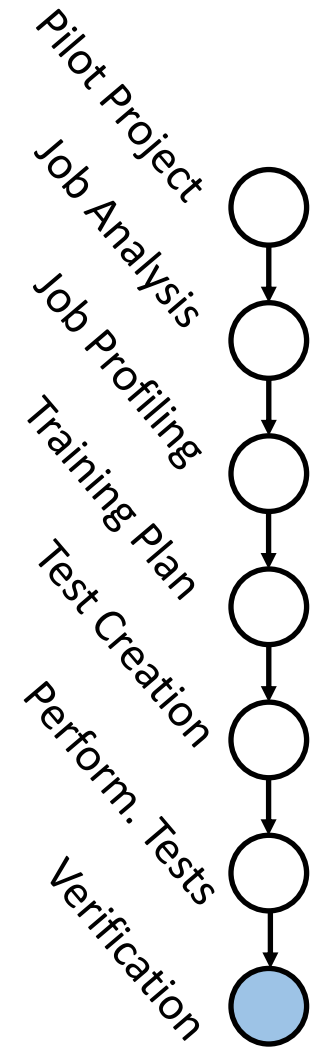
Comparing Performance of HT vs. PEMT search and cross-correlation (pilot project + professional translators test)

- **Students HT**
- **Students PEMT before training**
- Students PEMT after training
- Professional translators HT
- Professional translators PEMT before training
- Professional translators PEMT after training



Training Effectivity Verification

Spring 2017



Checklist for Training Materials Preparation

1. Pick-up a preferred domain (*e.g. audio/video user manuals*)
2. Find parallel texts (*e.g. manuals at a consumer electronics webpage*)
3. Align the texts (*SDL Trados Aligner™*)
4. Extract terminology – prepare terminology test (*manual*)
5. Prepare a comprehension texts
6. Text tagging (*see excel*)
 - identify types of errors
 - usability
7. Use a software for automated test production (*QuizCreator™ by Wondershare*)



Dear translator, here you can find 12 exercises which are focused at different skills and competencies vital for good post-editor of machine translation. You can pass through them step-by-step from No.1 to 12 or you can focus at specific ones to improve your weaknesses.



1. Decision making

As a posteditor you have only several seconds to distinguish if you use suggested MT or you rather delete it and translate text from scratch.



2. Morphological errors PEMT

Post-editing of segments with morphological errors only (discord in case, number, part of speech,...).



3. Typography errors PEMT

Post-editing of segments with typographical errors only (capital letters, punctuation,...).



4. Syntax errors PEMT

Post-editing of segments with syntactical errors only (word order).



5. Semantic errors PEMT

Post-editing of segments with errors in meaning only (wrong translation, missing translation)



6. Terminology errors PEMT

PE of segments with terminological errors only (usage of general voc. instead of domain specific).



7. Facts errors PEMT

Post-editing of segments with factual errors only (numbers, dates,...).



8. Error identification

Your task in this exercise will be to distinguish what kind of errors is represented in the segments.



9. PEMT with prompt

Your task is to change MT sentences according to given human translation. Here you are improving your PE skills and technical competence.



10. Free PEMT

Postediting of MT of user manual. This ex. imitate real-life postediting process and improve whole set of skills and competencies necessary for good professional post-editor.



11. Terminology

Mastering the terminology is vital not only for technical translator but also for post-editors. This exercise gives you opportunity to learn vocabulary used in user manual of audio-video appliances.



12. Topic comprehension

Reading materials. Perfect understanding of translating text is vital for good post-editors to prevent fatal errors done by MT engines.

EL TRADUCTOR COMO USUARIO AVANZADO DE INFORMÁTICA: HABILIDADES PARA LA TRADUCCIÓN AUTOMÁTICA

Amparo Alcina

Universitat Jaume I

<http://tecnolettra.uji.es/es/>

La era digital y globalizada en que estamos inmersos nos hace pensar que hoy en día cualquier persona es usuaria nata de la informática y se "lleva bien" con ella. Sin embargo, en el aula de traducción podemos observar numerosas deficiencias de conocimientos y habilidades informáticas de los estudiantes de traducción en el manejo de las diversas tecnologías. La traducción automática también se presenta en forma intuitiva para el gran público, pero el manejo que se exige al traductor de esta herramienta va mucho más allá de esta primera intuición. En este trabajo pretendemos caracterizar cuáles son las luces y sombras de los conocimientos, habilidades y actitudes en que deben formarse los traductores para el manejo de esta herramienta.

El traductor como usuario avanzado de informática

HABILIDADES PARA LA TRADUCCIÓN AUTOMÁTICA



Amparo Alcina, <alcina@uji.es>
Universitat Jaume I de Castelló

Planteamiento

- ▶ Como docente de las tecnologías de la traducción y terminología, me planteo qué debe aprender el estudiante que desea ser traductor y trabajar algún día como tal
 - ▶ conocimientos
 - ▶ destrezas
 - ▶ actitudes

Antecedentes

En relación con la traducción automática...

- ▶ QUÉ...?
- ▶ CÓMO...?
- ▶ CUÁNDO...?
- ▶ POR QUÉ...?
- ▶ QUIÉN...?



EXPERIENCIAS CON LA TRADUCCIÓN AUTOMÁTICA

Kingscott

- ▶ Aprender a teclear usando todos los dedos
- ▶ Aprender a usar todas las funciones del procesador de textos, no solo las básicas
- ▶ Dominar el manual de instrucciones del procesador de textos
- ▶ Usar abreviaturas de palabras que aparecen frecuentemente en el texto (función autotexto, etc.)

Weaver

Los sistemas de traducción interactiva piden al traductor realizar funciones como

- ▶ Interacción léxica
- ▶ Interacción sintáctica
- ▶ Interacción estilística

The manuals supplied by IBM contain the needed information.

information > dato // información

- Los datos almacenados en memoria se procesan inmediatamente
- La información almacenada en memoria se procesa inmediatamente

This demonstrates the high efficiency and sophistication of the Series 50 systems

IS THE TEXT SPEAKING OF: high sophistication > yes // no
Dependiendo de la respuesta:
...démontre le rendement et la complexité élevés...
...démontre le rendement élevé et la complexité...

Lawson

- ▶ El uso de TA en los servicios lingüísticos puede afectar la profesión.
- ▶ Se abren nuevos perfiles conforme los traductores aprendan a:
 - ▶ Editar los borradores de la TA
 - ▶ Construir diccionarios para la TA
 - ▶ Identificar las mejoras que son necesarias en el software

Ryan

- ▶ TA requiere la participación del traductor en cada etapa de su desarrollo y uso: desde
 - ▶ La selección del tipo de textos a ser traducidos,
 - ▶ La creación del diccionario
 - ▶ La creación de reglas gramaticales
 - ▶ Postedición
 - ▶ Presentación del producto final
- ▶ La implicación del traductor puede ser mínima, limitándose a proporcionar unos textos de ejemplo y un listado de palabras, a ser máxima y en ese caso también implicarse estrechamente en cada aspecto del desarrollo y convertirse en un co-desarrollador.

Ryan

Donde más se evidencia la implicación es el desarrollo de reglas de traducción:

- ▶ Sugerir áreas en las que el sistema necesita mejoras
- ▶ Destacar entre las mejoras aquellas que conllevan mayores pérdidas de tiempo al traductor o mayores cambios en el producto final.
 - ▶ Por ejemplo, un cambio en la forma en que se maneja determinadas formas de la construcción pasiva será más beneficiosa que resolver la ambigüedad gramatical de una palabra que finalmente es un ligero cambio en el producto final.
- ▶ El desarrollo de reglas para hacer efectivas las mejoras propuestas.
 - ▶ Mediante la observación de un problema detectado en cómo el sistema traduce un determinado aspecto, por ejemplo, la traducción de pasivas, y solucionarlo en la fase de posesición, puede formular de alguna manera qué podría hacer el sistema para cambiar la manera en que se traduce. El siguiente paso sería plantearlo y describirlo de una manera lógica.

Santangelo

Esta autora también destaca la posible participación del traductor en aspectos relacionados con el desarrollo del algoritmo.

- ▶ La postedición en el propio ordenador
 - ▶ Manejo de macroinstrucciones
 - ▶ Colaborar en el desarrollo de macroinstrucciones
 - ▶ Desarrollar la sensibilidad de detectar problemas que pueden no aparentar serlo
 - ▶ In fact > en efecto / en la práctica (p.135)
- ▶ Actualización del diccionario
 - ▶ Desarrollar la imaginación
 - ▶ Desarrollar la habilidad de pensar en términos binarios
 - ▶ Distinguir cuando cierto sentido de una palabra conviene que sea introducida en un microglosario (p. ej. Temático. p.ej. Smoking > fumar / ahumado (microglosario agricultura)
 - ▶ Anticipar problemas potenciales más que repararlos cuando ya han aparecido en un texto

Santangelo (2)

- ▶ Hacer recomendaciones sobre cómo mejorar el algoritmo de traducción
 - ▶ Diagnosticar el origen de los problemas: por ejemplo, si son causados por input incorrecto, palabras que no están en el diccionario, construcciones gramaticales incorrectas en el texto origen.
 - ▶ Comunicar al lingüista computacional los aspectos en que el algoritmo requiera mejoras

Además de la capacitación técnica necesaria para la ta, el traductor debe:

- ▶ Tener fe en el sistema de TA
- ▶ Estar comprometido en hacerlo funcionar
- ▶ Ser capaz de emocionarse en su desarrollo

McElhaney y Vasconcellos

Trabajar en la pantalla es mucho más eficiente. Aunque pueda parecer trivial:

- ▶ Buenas habilidades con el teclado
- ▶ Saber cómo mover el cursor rápidamente de un lugar a otro en el texto. Por ejemplo, la función de búsqueda, combinada con estrategias relacionadas, es mucho más rápido que la manipulación manual de las teclas de dirección (remite a Kingscott)

En resumen

- ▶ Hacer la mínima revisión; editar, no reescribir
- ▶ Trabajar con la pantalla, usar apoyo automático del procesador de textos siempre que sea posible
- ▶ Proporcionar feedback al sistema

Vasconcellos

Los traductores deben prepararse para hacer un compromiso a largo plazo para la construcción del sistema

Los traductores hacen su contribución a tres niveles (de acuerdo con Ryan y Santangelo):

- ▶ Agudizan sus propias habilidades y generan feedback para actualizar los diccionarios y mejorar el algoritmo
- ▶ Construyen el diccionario, y lo personalizan acorde con las necesidades propias
- ▶ Interactúan con los desarrolladores del sistema

Se requiere que los traductores tengan:

- ▶ Actitudes positivas
- ▶ Respuestas innovadoras
- ▶ Espíritu de superación



RECOPILACIÓN DE HABILIDADES

Habilidades cognitivas

- ▶ Buen conocimiento del sistema
- ▶ Ser capaz de deducir el origen de un mal resultado en la traducción
- ▶ Buen conocimiento de los efectos en el texto meta
- ▶ Conocer la jerga informática y comunicarse con ella

Destrezas

- ▶ Dominar el uso del teclado y otros periféricos del ordenador
- ▶ Dominar las funciones avanzadas de los programas informáticos. Ejemplo, el procesador de textos
- ▶ Elaborar pequeños programas (ej. Macroinstrucciones)
- ▶ Saber trabajar en equipo con informáticos
- ▶ Saber instalar programas, activar plugins

Destrezas (2)

- ▶ Desarrollar la habilidad de pensar en términos binarios
- ▶ Saber anticipar problemas potenciales (Detectar las palabras que pueden dar problemas en la traducción por ser ambiguas e implementar una solución adecuada: inserción en diccionario, incrementar los textos proporcionados al sistema...
- ▶ Diagnosticar el origen de los problemas
- ▶ Identificar los aspectos del software que necesitan mejorarse
- ▶ Destacar las mejoras cuya implementación representa mayor rentabilidad
- ▶ Desarrollar reglas para hacer efectivas las mejoras del sistema
- ▶ Saber plantear y describir las soluciones a los problemas del software de una manera lógica

Actitudes

- ▶ Autonomía para seguir aprendiendo por sí mismo. Por ejemplo, usando un manual de instrucciones del programa
- ▶ Iniciativa. Para aprender nuevas funciones de un programa, para implementar soluciones que ahorren tiempo, para poner en marcha soluciones imaginativas
- ▶ Tener fe en el software
- ▶ Estar comprometido en hacer funcionar el software
- ▶ Ser capaz de alegrarse con la mejora del software



CONCLUSIÓN

Conclusión

- ▶ La traducción automática y las tecnologías de la traducción en general exigen habilidades informáticas de nivel avanzado que implican que el traductor se involucre en algún nivel en el desarrollo del sistema
- ▶ Es necesario incorporar mayores conocimientos de informática en la formación de traductores
- ▶ Es necesario incorporar más práctica con ordenadores en la formación de traductores
- ▶ El estudiante de traducción debe prepararse para ver en el ordenador un aliado
- ▶ La imagen que ofrece la profesión del traductor debería incorporar la fuerte presencia de las tecnologías

Bibliografía

- Kingscott, G. 2008, "Translator Strategies for Getting the Most. Out of Word Processing" , en M. Vasconcellos (ed.) 2008, pags. 14-18.
- Lawson, V. 2008, "A Translator Map of Machine Translation" , en M. Vasconcellos (ed.) 2008 , pags. 105-115.
- McElhaney, T. and M. Vasconcellos, 2008, "The Translator and the Postediting Experience", en M. Vasconcellos (ed.) 2008 , pags. 140-148.
- Ryan, 2008, "The Role of the Translator in Making an MT System Work Perspective of a Developer", en M. Vasconcellos (ed.) 2008 , pags.127-132.
- Santangelo, "Making an MT System Work: Perspectives of a Translator" , en M. Vasconcellos (ed.) 2008 , pags.133-139.
- Vasconcellos, 2008, Factors in the Evaluation of MT: Formal vs. Functional Approaches, en M. Vasconcellos (ed.) 2008 , pags.203-213.
- Vasconcellos, M. (ed) 2008, *Technology as Translation Strategy*. John Benjamins.
- Weaver, 2008, Two Aspects of Interactive Machine Translation", en M. Vasconcellos (ed.) 2008 , pags.116-123.



Gracias por su atención

EXPLOITATION OF A COMPARABLE CORPUS OF USER-GENERATED CONTENT FOR POST-EDITING STRATEGIES

Miguel Ángel Candel-Mora

Universitat Politècnica de València

<http://www.upv.es/entidades/DLA>

With the Web 2.0 and the active participation of users, online consumer-generated reviews have become a clear reference in purchasing decision-making processes, and on occasions sometimes exceed the reliability and authenticity of advertising campaigns from manufacturers.

User reviews may appear in different formats and with different structures (Vázquez, 2014): as evaluations of a product, in the form of dialogue in a forum, or in the case of this study, as unstructured free text for the evaluation of a tourism product (hotel reviews), in which, in addition to specific linguistic features, other aspects such as the reviewers' reliability, naturalness, or credibility play a key role.

Hotel reviews in particular have been studied extensively (Schemmann, 2011) and research has clearly identified its structure and revealed that opinions are not only conveyed through linguistic resources, but there are other genre-specific features such as intertextuality, the profile of the reviewer, and paralinguistic elements that contribute to the reliability and credibility of consumer reviews (Pollach, 2006; Vázquez, 2012).

From the point of view of post-editing, and keeping in mind the purpose of the target text (O'Brien, 2005), work to improve machine translation would basically consist in repairing and accommodating the MT output - in varying degrees - to the appropriate target language linguistic conventions.

Thus, the research question that motivates this paper is that in the case of user-generated reviews in the domain of tourism, the message is not only transmitted through linguistic resources but there are other elements or textual artifacts that should be taken into consideration in the post-editing strategy, in addition to relevant grammar and stylistic PE guidelines (Babych, 2014; Vilar et al., 2006).

Several studies have already confirmed that there are no universal guidelines for post-editing (Allen, 2003; TAUS, 2010), and each genre requires specific quality rating scales. Thus, this work highlights the need to pay special attention to the textual conventions during any post-editing strategy in addition to identifying error patterns common to most

post-editing guidelines. More specifically, the objective of this work is to compare textual characteristics of user reviews originally written in English and in Spanish from data derived from a corpus-based approach analysis that serve as standards in MT output post-editing tasks.

Furthermore, this objective is further expanded with two subobjectives:

- 1- to extract MT-related data to elaborate training material for future pos-editors, and
- 2- to reuse the translation knowledge extracted to add bilingual dictionaries to the parallel corpus of the MT system to improve its performance.

Exploitation of a comparable corpus of user-generated content for post-editing strategies

Miguel Ángel Candel-Mora
Dept. Applied Linguistics
Universitat Politècnica de València
mcandel@upv.es

- 1. Introduction**
- 2. Objective**
- 3. Methodology**
- 4. Results and discussion**
- 5. Conclusions and future work**

The facts

CGR

Web 2.0 & active participation of users

online consumer-generated reviews = referent in purchasing decision-making processes

The facts

CGR

TripAdvisor Content:

385 million reviews and opinions from travelers around the world.

More than 6.6 million businesses and properties in 135,000 destinations, including:

1 million+ hotels, B&Bs, and specialty lodging

815,000 vacation rentals

4.1 million restaurants

690,000 attractions and experiences

More than 70 million candid traveler photos.

More than 255 new contributions are posted every minute.

110 million marketable members worldwide.

On average, nearly 2,600 new topics are posted every day to the TripAdvisor forums.²²²

Tourism 2.0



“seven in every ten Internet users worldwide trust consumer opinions and peer recommendations posted online”

(Schemmann, 2011)

The facts

MT & CGR

review platforms use machine translation systems to immediately make that review available to as many users as possible in different languages

raw machine translations without further processing or revision

The facts

MT & CGR

[opinions are not only conveyed through language]

For MT purposes:

genre-specific features

- **intertextuality,**
- **the profile of the reviewer,**
- **paralinguistic elements**

**contribute to the reliability and credibility of
consumer reviews**

Postediting research

Research on MT post-editing has been approached from different points of view

- **Quality**
- **Evaluation guidelines**
- **Productivity gain**
- **Cognitive effort**
- **Acceptability of MT output**

INITIAL PREMISE



No universal post-editing guidelines

New genre = new guidelines

2. Objective

to compare genre characteristics of user generated reviews from data from a corpus-based approach analysis that serve as standards/guidelines for MT output post-editing tasks

2. Objective

- 1- to extract MT-related data to elaborate training material for future post-editors, and**
- 2- to reuse and add translation knowledge to the parallel corpus of the MT system to improve performance.**

ProjecTA



research project funded by the Spanish Ministry of Economy and Competitiveness, aimed at exploring the effects of the implementation of MT-related services on the professional profile of translators.

CONSUMER GENERATED CONTENT AS NEW DIGITAL GENRE

Terminology:

“electronic word of mouth”

“eWOW”

“online consumer reviews”

“user generated product reviews”

“product reviews”

“user opinions”

evaluation of users posted on a travel review site on their experience.

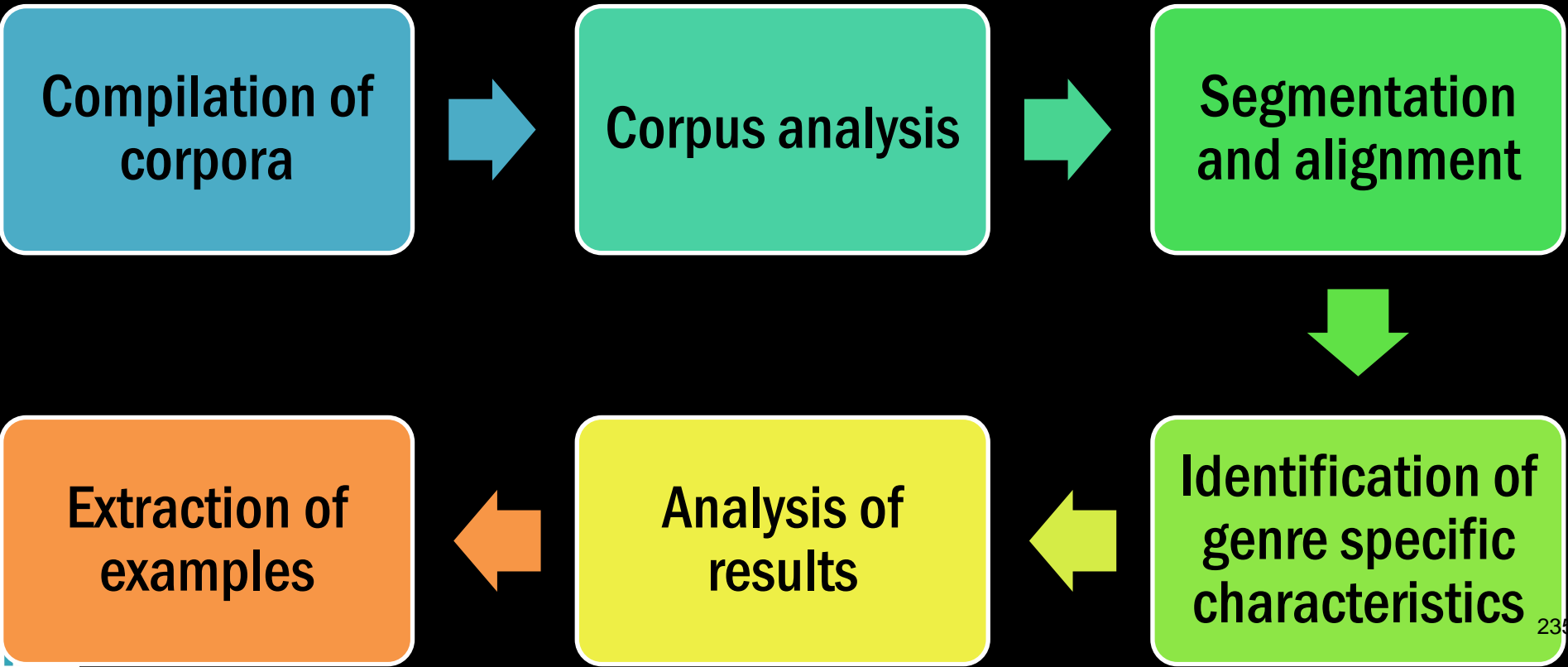
CGR RESEARCH LINES

- potential roles of product reviews in the decision-making process (Ricci & Wietsma, 2006)
- involvement of reviewers (Vásquez, 2014)
- characterization of online reviews (Shemmann, 2011),
- improvement of review websites (Pollach, 2006).

evaluations of a product,
in the form of dialogue in a forum,
unstructured free text for the evaluation of a
tourism product

1. chronological sequence of events
2. reported speech, story prefaces, deictic shifts
3. humor, detail and personal experience
4. nonlinguistic cues (punctuation...)
5. intertextuality – reference to previous comments
6. strategies designed to **compensate the impersonality** of written discourse

3. Methodology



3. Statistics

text file	corpuscgr_t3l_en. txt	corpuscgr_t3l_ mt-es.txt	corpuscgr_t3l_e s.txt
tokens (running words) in text	130.439	139.160	80.939
types (distinct words)	6796	8462	7145
type/token ratio (TTR)	5,265645504	6,14163065	8,914869308
standardised TTR	41,21007156	42,17956924	42,25
standardised TTR std.dev.	57,97714233	57,24690628	55,74359131

The data

Average # of words per review

Corpus CGR EN	Corpus CGR MT-ES	Corpus CGR ES
128,05 words	137,02 words	78.85 words

The data

Shortest review in Corpus CGR EN	Shortest review in Corpus CGR ES	Shortest review in Corpus CGR MT-ES
26 words	8 words	30 words

The data

Sample reviews

“Excelente calida/precio. Ubicacion inmejorable. Limpio y sin ruidos.”

Multiple visits reinforce the quality & service at this boutique hotel, good food, excellent service.
Location is perfect for both city and west end, lovely neighbourhood.

The data

Longest review in Corpus CGR EN	Longest review in Corpus CGR ES	Longest review in Corpus CGR MT-ES
1196 words	980 words	1284 words

The data: average

50% reviews in ES 8-50 words

50% reviews in EN 20-90 words

The data sample reviews & length

75 words

This hotel was in an ideal location its about 6 mins walk from West Kensington tube station, We were greeted as we checked in which was quick ,the room was on the 2 floor it had all we needed. The breakfast was very good with a good choice to help to set up for the day .Around the area there are num supermarkets and pubs would use this hotel again.to get the best price deal with direct.

The data sample reviews & length

50 words

Muy bien ubicado a apenas dos calles de la Gran Vía, la zona por la noche se llena de prostitutas pero no por eso se vuelve insegura. Personal amable y siempre dispuesto. Habitación cómoda y limpia aunque bastante modesta. Desayuno sencillo pero correcto. Servicio razonable con lo que uno paga...

4. Results and discussion

PE Scales [potential problems]

vocabulary / terminology

syntax

ambiguity

Genre-specific error patterns [naturalness]

acknowledgements / thanks

advice

intertextuality and reference to other opinions

demonstrate authority/ reliability

frequent multiword expressions

usual content words

PE scales vocabulary

EN	MT	ES
convenient	conveniente	Conveniente , cómodo, práctico, bien situado, adecuado
amenities	artículos de aseo (Amenities) 0	amenities
busy	ocupado	ajetreado, concurrido, transitado, animado

PE scales

syntax

Loved it & hope to return soon.

Me encantó y esperamos volver pronto.

when on business

cuando por negocios

Well located and will definitely stay here again when in the area.

Bien situado y sin duda me alojaría aquí de nuevo cuando en la zona.

Who needs a pool when you are staying at the Montague when in London?!?

¿Quién necesita una piscina cuando se aloja en el Montague cuando en Londres?!?

PE scales

word agreement

ST: ... a stay here is not **cheap**.

MT: ...una estancia aquí no es **barato**.

ST: We had two rooms and both were **perfect** in every way.

MT: Teníamos dos habitaciones y ambas eran **perfecto** en todos los sentidos.

ST: The hotel also **booked** theatre tickets for me.

MT: El hotel también **reservamos** billetes de teatro para mí.

Error Pattern Identification in Consumer Reviews: Mistranslations

ST: I had selected a few **plays**...

MT: Había seleccionado **algunos** ~~juega~~ obras de teatro...

ST: ...they know what you have **had**.

MT: ...saben lo que hemos ~~tenido~~-tomado

ST: My phone only **charges** with that charger

MT: mi teléfono sólo ~~cobra~~ carga con ese cargador

Genre specific features

naturally-occurring stylistic
preferences – Corpus CGR-ES

evaluations

bien ubicado, cómodo, con detalles agradables y la limpieza de 10!

Buena relación calidad precio.

como en casa

Está muy bien localizado

hotel muy recomendable

La atención del personal ha sido excelente

la mejor opción

la relación calidad

La ubicación del hotel está muy bien

La ubicación es la mejor,

Pero sin duda lo mejor es la atención del personal.

Un 10 para todo el personal del hotel.

Un hotel urbano super chulo

un tres estrellas

vale la pena

Genre specific features

naturally-occurring stylistic preferences

acknowledgements / thanks

Gracias

Mil gracias por TODO de nuevo!!

Gracias por todo, volveremos.

Doy las gracias al personal de recepción por su amabilidad y profesionalidad.

Gracias nuevamente a todo el personal del hotel por que te hacen sentirte como en casa.

Muchas gracias a todo el personal

Muchísimas gracias al equipo especialmente Clara y MJosé por ayudarnos a organizar todo."

Sin duda repetiré en otro momento, muchas gracias por todo."

Solo puedo decir gracias y recomendarlo SIEMPRE "

Cuando vuelva a Madrid , espero hospedarme en este hotel
hotel muy recomendable nosotros con casi total seguridad
volveremos

Repetiré cuando vuelva a
sin duda lo recomiendo y volvería a hospedarme ahí
sin duda volveré

Genre specific features

intertextuality and reference to other opinions

"Me quede en este hotel **debido a las buenas críticas** que había leído y desde luego razón tenían.

No alojamos en este lugar **por las críticas de Tripadvisor** y la verdad es que estamos muy contentos de haberlo hecho.

Después de **tantas críticas buenas** y las fotos tan bonitas, nos quedamos con cara de idiotas.

Primera vez que reservamos sin recomendación de terceros, **solamente por trip**.

"Guiado **por las opiniones de Tripadvisor** seleccionamos este hotel de 3 estrellas.

Teníamos unas expectativas **,por lo que habíamos leído en Tripadvisor**, que no se cumplieron.

Tenía **muchas dudas porque había leído críticas muy malas** sobre lo viejas que están las habitaciones, pero también otras buenas, pero por necesidad, me arriesgué.

Genre specific features

frequent multi-word expressions – CORPUS CGR-ES

amabilidad del personal
buen servicio de
buena atención del
completo y variado
con todas las comodidades
con todo lo necesario
con vistas a la
cuenta con un
daba a un patio interior
dandole un plus que
de cuatro estrellas
de lo mejor que
de muy buena
dieron una habitación
El hotel se encuentra
El personal es muy atento
el trato recibido

en cuanto al
es una maravilla
está muy bien situado
está muy bien ubicado
Habitaciones espaciosas
hotel perfecto para
inmediaciones
la única pega es
Merece la pena
muy bien comunicado
para mi gusto
por lo demás
por motivos de
sin lugar a dudas
Volveremos en nuestra próxima visita
Hotel muy bien situado

Genre specific features
naturally-occurring stylistic
preferences

frequent multi-word expressions

Corpus CGR EN	Corpus CGR MT-ES	Corpus CGR ES
home away from home	hogar lejos de casa	como en casa
good value for money	relación calidad-precio	relación calidad/precio Relacion calidad precio

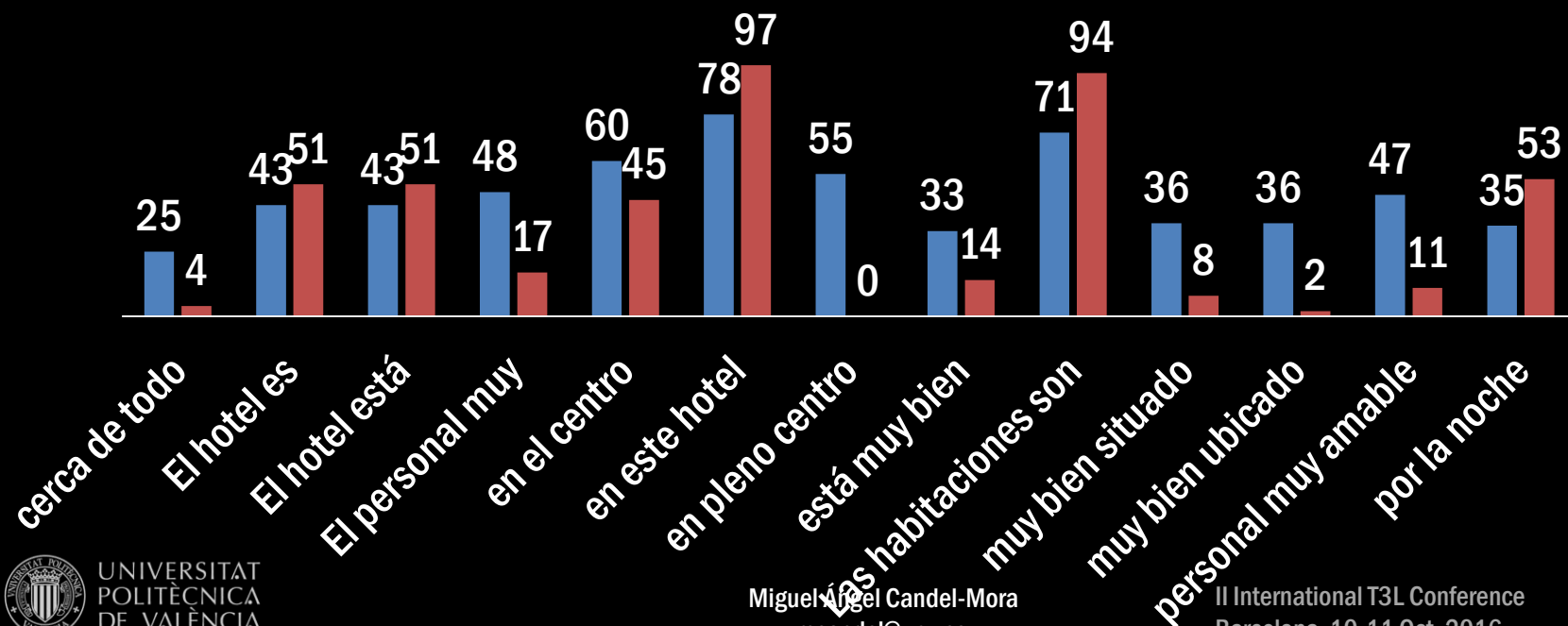
Genre specific features usual content words

9 HOTEL	1248	10 HOTEL	1782
22 HABITACIÓN	548	13 ROOM	1420
28 HABITACIONES	429	24 STAFF	883
30 PERSONAL	410	34 STAY	619
33 DESAYUNO	352	40 BREAKFAST	526
40 UBICACIÓN	247	44 LOCATION	412
46 PRECIO	209	47 ROOMS	391
47 SERVICIO	207	54 NIGHT	335
50 CENTRO	196	58 SERVICE	304
52 BAÑO	187	68 BED	257
54 CALIDAD	183	83 DAY	206
55 NOCHE	179	84 WALK	200
58 RECEPCIÓN	167	88 PLACE	194
64 METRO	154	90 BATHROOM	192
67 TERRAZA	140	93 RECEPTION	185
68 ZONA	140	94 SHOWER	185
70 ATENCIÓN	135	95 STATION	181
71 CAMA	135	98 FOOD	175
72 PUERTA	135	102 TEA	165
73 TRATO	131	116 BAR	151
81 LIMPIEZA	108	118 NIGHTS	149
86 CALLE	106	123 HOTELS	139
108 VISTAS	86	124 AREA	138
113 CAMAS	80	131 TUBE	134
115 ESTANCIA	79	134 FLOOR	130
116 HABITACION	79	138 EXPERIENCE	127
122 DECORACIÓN	75	139 BIRTHDAY	125
123 DUCHA	75	146 COFFEE	118

Genre specific features

frequent multiword expressions

■ occurrences in ES ■ occurrences in MT



Genre specific features

Adjectives not in CGR-MT-ES

ADJ in corpus CGR_ES			ADJ in corpus CGR MT-ES		
Position in word list	ADJ	occurrences	Position in word list	ADJ	occurrences
38	BUENA	257	56	GRAN	299
39	EXCELENTE	256	58	AMABLE	290
57	AMABLE	167	62	NUEVO	265
59	BUEN	166	68	EXCELENTE	247
60	GRAN	164	71	MEJOR	228
65	POCO	153	75	AGRADABLE	219
75	MEJOR	118	76	POCO	218
87	MUCHO	106	84	SERVICIAL	195
88	LIMPIO	105	86	BUENA	190
95	RECOMENDABLE	97	104	PEQUEÑA	152

Frequent words in hotel reviews & ambiguity

- bar (establishment / counter / candy)
- play (sport / theater)
- ticket (train / theater)
- glass (receptacle / material)
- common English verbs that have two forms in Spanish:
to be, to have, to miss,

Genre specific features

Terminology

caja de seguridad
cama de matrimonio
desayuno buffet
el personal del hotel
el wifi funciona
hacer el check-in
hacer la reserva
la reserva en
nos asignaron una habitación
servicio de desayuno
Ubicado en el centro

5. Conclusions and future work

The nature of the reviews conditions to a large extent the need to propose specific post-editing guidelines

translation quality scales include error annotation and calculation of proportion of errors with the total amount of words in the translated text

in the case of consumer reviews, consisting of free text of reduced dimensions, the error proportion would be higher and low quality translation would be more noticeable.

5. Conclusions and future work

genre-specific features are essential for naturalness, reliability and credibility of consumer reviews

naturally-occurring preferences

Among the MT-related data extracted from comparable corpora are collections of translation equivalent fragments of text, such as terminological expressions, frequent multi-word expressions, or usual content words and collocations

Future work

- Change the perspective to ES-EN
- application of methodology to other types of texts susceptible to be processed by machine translation systems
- with other combinations of languages

Exploitation of a comparable corpus of user-generated content for post-editing strategies

Thanks!

Miguel Ángel Candel-Mora
Dept. Applied Linguistics
Universitat Politècnica de València
mcandel@upv.es

MACHINE TRANSLATION IN SPANISH LSPs

Olga Torres-Hostench

Marisa Presas

Pilar Cid

Universitat Autònoma de Barcelona

projecta.tradumatica.net

In this presentation we will discuss with the audience the results of a report carried out on the use of Machine Translation (MT) and MT Post-editing (PE) in Spanish Translation Service Providers in 2015. The report is part of a research project called ProjecTA, funded by the Spanish Ministry of Economy and Competitiveness (Ref. FFI2013-46041-R) and aimed at analysing the flow of MT+PE work in the professional translation sector in Spain. Quantitative data were collected through an online survey, which was reviewed by experts in Statistics and TSP representatives. The survey was sent individually to 189 Spanish TSPs during January and February 2015 and 55 surveys received were valid, which corresponds to the 29.5% of the sample. The survey consisted in 17 questions, distributed in three areas: (a) basic identification information from the TSP (location, number of employees, annual turnover, etc.); (b) Services and specialties offered by the TSP, most common source and target languages, types of customers, etc.; (c) use of MT and PE in the company (use of MT; reasons not to use MT; percentage of MT use in the company projects; type of MT system; ownership of the MT system; translators' acceptance of PE work; percentage of PE use in the company projects). . The quantitative results show that 47.3% of Spanish language service providers use MT and that 45.5% of these use MT in only 10% of their total projects. Furthermore, qualitative data were also obtained via telephone interviews, personal interviews with experts and a focus group session in second quarter 2015. The qualitative results reveal that the decision to implement an MT system depends on multiple factors: the business's economic capacity, technological capacity and the knowledge and attitude of the business's human resources.



Machine Translation in Spanish LSPs

Olga Torres-Hostench, Marisa Presas, Pilar Cid (coord.)

2nd International T3L Conference: Tradumatics, Translation Technologies & Localisation
"Translators and machine translation"

9th International Conference on Translation,
Department of Translation, Interpreting and East Asian Studies
10-11 October 2016 - Universitat Autònoma de Barcelona



Aims of the presentation



1. Disseminate the report: El uso de traducción automática y posesición en las empresas de servicios lingüísticos españolas : informe de investigación ProjectA 2015
2. Debate with the audience some of the controversies on MT derived from the report.



1. Disseminate the report



Download the report for free from the DDD (digital document repository of Universitat Autònoma de Barcelona) at:

<http://ddd.uab.cat/record/148361>



Controversies



Controversy 1



Why using MT (and even talking about it)
is still a taboo for many Spanish LSPs?

Phone conversations:

“We work with translators from flesh and bones”

“We do everything with our head”

“We do everthing with traditional methods”

<http://ddd.uab.cat/record/148361>



Controversy 2



Why some LSPs use MT, but do not tell their clients (the same with freelance translators)?

<http://ddd.uab.cat/record/148361>



Controversy 3



Why some (most?) Spanish LSPs do not make an effort to keep pace with MT technologies?

<http://ddd.uab.cat/record/148361>



Brief summary of the report



Methodology

Most significant quantitative results

Most significant qualitative results



Methodology



Aim

- Measure the implementation of MT in Spanish LSPs
 - To what extent do they use MT and PE?
 - Which is the profile of companies using MT and PE?

Methodology

- Online survey was sent to 187 companies (55 answers)
 - 17 questions on
 - Companies profile
 - Use of MT and PE
- Focus group session with experts



Most significant quantitative results



Q: Is MT used in your workflow?

A: **47.3% use MT, 52.7% do not use it.**

Q: If MT is not used, why?

A: **'We do not trust MT' (35.6%); 'Clients do not ask for MT' (33,3%); 'Translators reject it' (20%).**

Q: 'Which is the percentage of projects for which you use MT?'

A: **0-10% (45.5%) ... +90% (3.6%)**

Q: Does your company have a customized MT engine?

A: **42.3% do own a customized MT engine; 19.2% do not have it; 38.5% do not answer.**



Most significant qualitative results



Some companies (and some translators) use MT without admitting it.

When MT is merged with TM, translators may not be aware that they are using MT.

Free MT engines do not protect confidentiality.

Implementation costs of customized MT engine are high.

Lack of technological capacity of companies.

Lack of MT training, prejudices from translators.



Qualitative results: Obstacles for MT implementation



Linguistic combinations

Subject area of assignments

Tight deadlines

Different client profiles and specific demands

Company reluctance to changes in workflow.

Decrease in productivity during preparation of the engine and during training

Acceptance by translators



Qualitative results: MT challenges for LSPs



Invest in a customized MT system which guarantees confidentiality and translation quality

Encourage PEMT and MT training for translators working for them

Overcome fears and reluctance towards MT

What do YOU think?

We would like to hear your thoughts

<http://ddd.uab.cat/record/148361>

ON THE CORRECTNESS OF MACHINE TRANSLATION: AN ENGLISH-ITALIAN CASE STUDY

Giovanni Cerasani

Binari Sonori

www.binarisonori.it

In order to understand the present capability of machine translation, I drew inspiration from the essay “On the correctness of machine translation: A machine translation post-editing task” to make a study of five texts, different from each other in both difficulty and subject. They were analysed between May and July 2015 through the Google Translate (statistical) and SYSTRANet (hybrid) systems, as my intention was to verify the reliability of free MT, which is more likely to be used by common web users.

Unlike the quoted study, my analysis on the English-Italian MT output was carried out with the source texts constantly at hand. Therefore I adopted different evaluation methods from the 1-to-5 scale originally used for both meaning and grammar. Since most of the sentences were long and syntactically complex, and since I wanted to focus on the recognition of a “core meaning”, I chose a 1-to-4 scale in order to evaluate the semantic accuracy:

- 1 – The core meaning is obscure or distorted;
- 2 – The core meaning is not very clear;
- 3 – The core meaning is clear;
- 4 – The whole sentence is clear.

As far as grammar is concerned, in my opinion neither the quality nor the quantity of errors can be mathematically established. Not only do the various error types not weigh the same, but a single type may not be equally difficult to understand and correct, according to the context. That’s the reason why I chose a 1-to-3 scale:

- 1 – The average error quantity and quality requires massive post-editing;
- 2 – The average error quantity and quality requires light post-editing;

- 3 – The sentence is correct.

We can extract some interesting data from the 376 sentences analysed (188 for each system):

- 184 sentences (49%) are completely clear. 28 of these (7%) are grammatically correct;
- 148 sentences (39%) are ungrammatical, and 41 of these (11%) are obscure.

The most common semantic error is the incorrect translation of polysemous words. English terms may occasionally be kept: sometimes they are out-of-context anglicisms generally accepted in the Italian language today; some others they are non-standard English words which are, thus, not identified by the MT systems themselves. Multiword expressions may also be tricky and, as a consequence of the individual translation of the lexemes, the overall meaning is often lost.

As for grammar, it may be the case that verbs are not translated in the most suitable tense or mood: inconsistencies can sometimes be found in a single sentence. Phrasal verbs are a source of misunderstandings both from a semantic (i.e. the “get” phrasal verbs) and a syntactic perspective (when prepositions are taken separately). Word order is mostly retained: it is certainly positive with SVO and brief sentences, but it is not when the latter are longer and have a more complex syntax. Gender and number agreement are not always guaranteed.

On the correctness of machine translation: an English-Italian case study

Giovanni Cerasani

**II International T3L Conference: Tradumatics, Translation Technologies & Localisation
"Translators and Machine Translation"**

**IX International Conference on Translation
Department of Translation, Interpreting and East Asian Studies**

Free MT reliability

- o Two systems used:
 - o Google Translate (statistical, 500M+ users)
 - o SYSTRANet (hybrid, ~50 years in the field)
- o 5 texts translated EN>IT (4 articles, 1 fiction; 188 sentences) in May-July 2015
- o No test subjects

Semantic evaluation scale

1. The *core meaning* is obscure or distorted
2. The *core meaning* is not very clear
(contextualisation required)
3. The *core meaning* is clear (but
the secondary, if any, is not)
4. The whole sentence is clear (despite
grammar mistakes)

Grammatical evaluation scale

1. The average error quantity and quality requires massive post-editing
2. The average error quantity and quality requires light post-editing
3. The sentence is correct

Texts analysed

- A. «insect» text, zoology. 28 sentences, avg WPS 23.6
- B. «telecom» text, technology. 32 sentences, avg WPS 21
- C. «conscience» text, philosophy/bioethics. 37 sentences, avg WPS 18.4
- D. «Faustus» text, narrative. 65 sentences, avg WPS 10.4
- E. «Holmes» text, film review. 26 sentences, avg WPS 26

*19 International T3L Conference: "Translators and Machine Translation"
Universitat Autònoma de Barcelona, 10-11 October 2016*

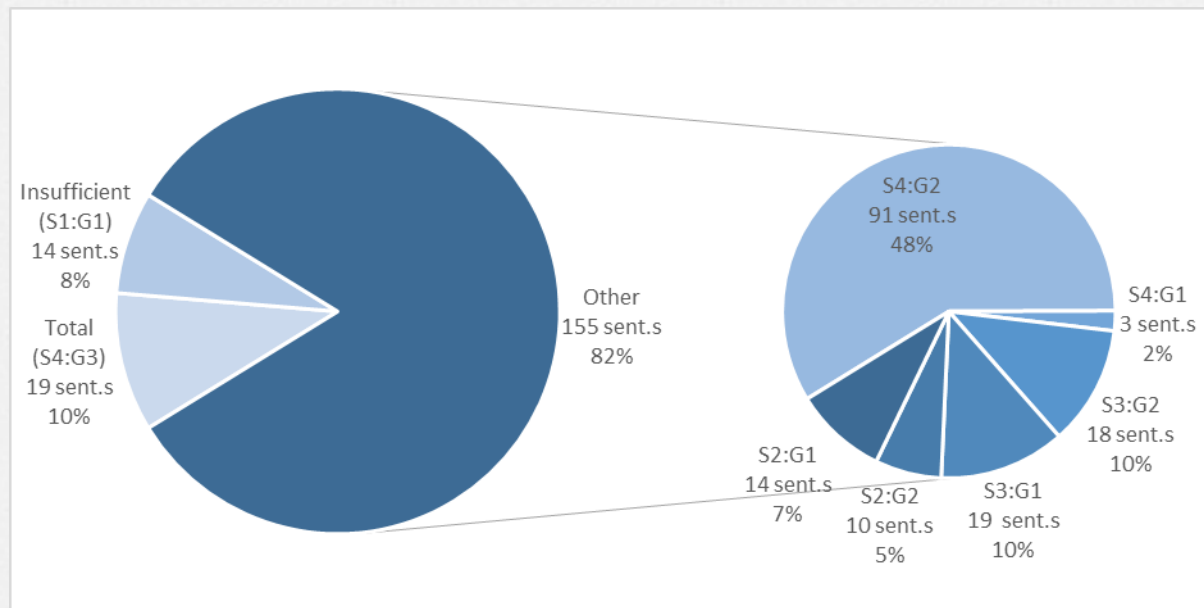
GT/SYSTRANet comparison

PROS	<i>Google Translate</i>	<i>SYSTRANet</i>
Sentences w/ exact meaning(s)	113 (60%)	71 (38%)
Publishable sentences	19 (10%)	9 (5%)

CONS	<i>Google Translate</i>	<i>SYSTRANet</i>
Ungrammatical sentences	50 (27%)	98 (51%)
Sentences w/o meaning	14 (8%)	27 (14%)

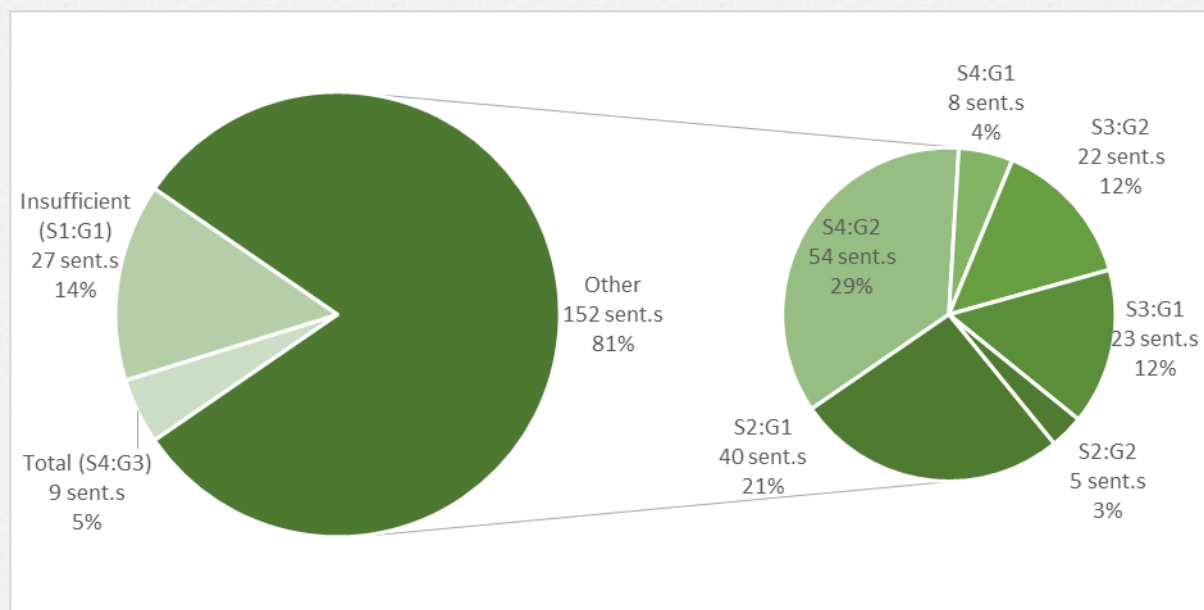
19 International T3L Conference: "Translators and Machine Translation"
Universitat Autònoma de Barcelona, 10-11 October 2016

Google Translate correctness



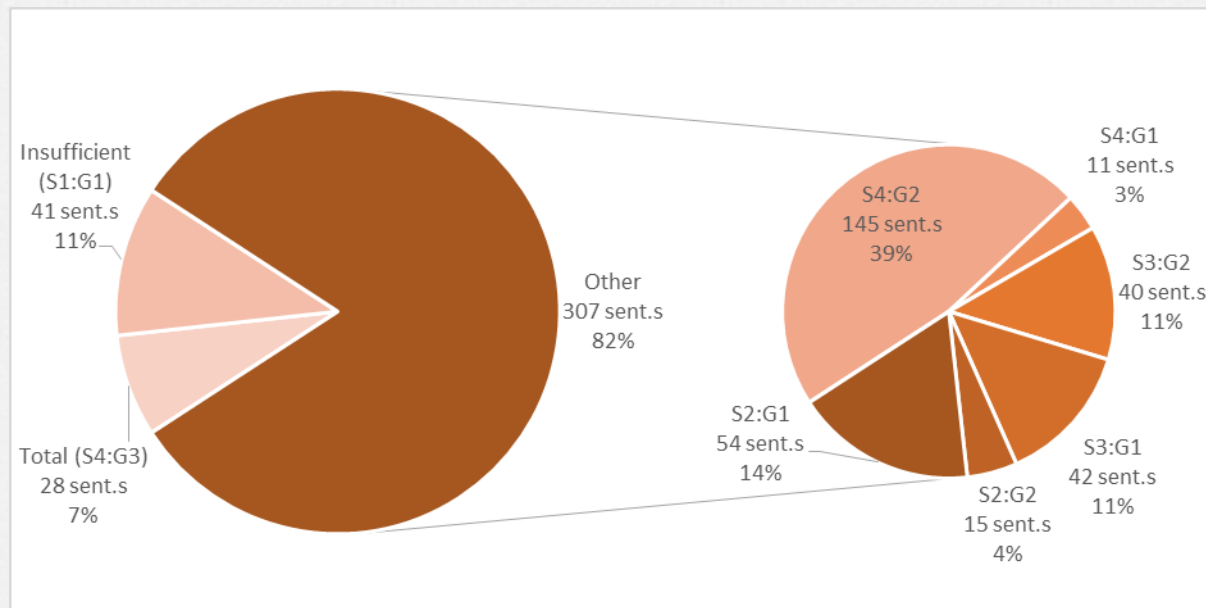
11 International T3L Conference: "Translators and Machine Translation"
Universitat Autònoma de Barcelona, 10-11 October 2016

SYSTRANet correctness



11 International T3L Conference: "Translators and Machine Translation"
Universitat Autònoma de Barcelona, 10-11 October 2016

Overall correctness



11 International T3L Conference: "Translators and Machine Translation"
Universitat Autònoma de Barcelona, 10-11 October 2016

Overall salient data

- o 184 (49%) sentences are clear
 - o 28 (7%) are also grammatically correct
 - o 7 pair of sentences shared: A1, A12, A19, C1, C4, C19, D31
- o 148 sentences (39%) are ungrammatical
 - o 41 (11%) are also obscure
 - o 8 pair of sentences shared: A5, A26, B1, B9, B10, D8, D22, E16

Semantic errors: Polysemous words

- o D8: His studies now seemed **dull** to him and he did not know what he should do.
 - o (GT) I suoi studi ora sembrava **sordo** a lui e lui non sapeva che cosa *avrebbe dovuto* fare
 - o (SN) I suoi studi ora sono sembrato **smussati** a lui e non *ha conosciuto* che cosa *dovrebbe* fare
- o sordo= muffled (sound)
- o smussato= not sharp (blade)
- o dull= boring > monotono, noioso

Semantic errors: English words

o Out-of-context Anglicisms

- o «bug» instead of «*insetto*» in B text (IT term)

o Unidentified non-standard English words

- o «critter» in A4 («creature» > «*creatura*»)
- o «trad» in E7 (abbreviation of «traditional» > «*tradizionale*»)

Semantic errors: Multiword expressions

- o B1: **Home truths** about telecoms
 - o (GT) *Home truths* circa telecomunicazioni
 - o (SN) *Verità domestiche* circa le telecomunicazioni
- o home= casa (n.), domestico (adj.)
- o truth(s)= verità
- o home truths= scomode verità (inconvenient truths)

Semantic/syntactic errors: Phrasal verbs

- o D41: Dr Faustus had made up his mind > Il Dott. Faustus ***aveva composto la sua mente***
 - o To make up one's mind= *decidersi, prendere una decisione*
- o D35: He **reached out** for another book > ***Ha raggiunto fuori*** per un altro libro
 - o To reach out= *allungare le braccia, allungarsi, protendersi* (formal)

Syntactic errors:

Wrong tense/mood

- o A3, past simple:
 - o Took > *ha preso* (*passato prossimo*)
 - o Returned > *tornò* (*passato remoto*)
- o A4: Despite working > *Nonostante che lavora* (*Nonostante [egli] lavori*)
 - o «*nonostante*» requires the *congiuntivo* mood
- o B24: «finding» treated as a gerund, «*che trova*», rather than a noun, «*scoperta*»

Syntactic errors:

Retained WO/GN agreement

A2: As **collections manager** at London's Natural History Museum, Max Barclay has traveled the world in search of rare and previously undiscovered insects > Come **collezioni direttore** (curatore delle collezioni) al Museo di Storia Naturale di Londra, Max Barclay ha viaggiato per il mondo in cerca di insetti rari e finora sconosciuto (sconosciuti).

Google (re-)Translate

- o September 26°, 2016
 - o 78 sentences (42%) as 2015
 - o 90 sentences (48%) same score as 2015 with slight differences
 - o 10 sentences (5%) better than 2015
 - o 10 sentences (5%) worse than 2015
- o Main difference: polysemous words translated differently (and, sometimes, wrongly)

*11 International T3L Conference: "Translators and Machine Translation"
Universitat Autònoma de Barcelona, 10-11 October 2016*

Which future for MT?

A future of collaboration between humans and machines in order to knock down linguistic and cultural barriers.

Hoping that this, like the Babel Fish imagined by Douglas Adams, will not cause

“more and bloodier wars than anything else in the history of creation”.

(Douglas Adams,
The Hitchhiker's Guide to the Galaxy)

19 International TSL Conference: "Translators and Machine Translation"
Universitat Autònoma de Barcelona, 10-11 October 2016

Bibliography

- “A scientific detective story”. 2008. [ENGLISH WRITTEN PART PITKÄ
OPPIMÄÄRÄ LÄNG LÄROKURS 17.3.2010](#), 1-2
- Adams, Douglas. 1979. *The Hitchhiker’s Guide to the Galaxy*. London: Pan Books
- Bradshaw, Peter. 2015. “Mr Holmes review – Ian McKellen gets more fascinating with age”. [The Guardian](#)
- “Home truths about telecoms”. 2007. [The Economist](#)
- Koponen M., L. Salmi. 2015. “On the correctness of machine translation: A machine translation post-editing task”. [The Journal of Specialised Translation 23](#), 118-136
- Marlowe, Christopher. 2002. *The Tragedy of Dr Faustus*. Canterbury: Black Cat Publishing
- “Translate where you need it: in any app, offline, and wherever you see Chinese”. 2016. [Google Translate Blog](#)
- “What is consciousness?”. 2006. Objective IELTS Workbook with Answers, 35. CUP

*11 International T3L Conference: “Translators and Machine Translation”
Universitat Autònoma de Barcelona, 10-11 October 2016*

LOS ERRORES DE LA TRADUCCIÓN AUTOMÁTICA CHINO > ESPAÑOL

Hong Zhang

Universitat Autònoma de Barcelona

www.uab.cat

El presente trabajo tiene como prioridad principal ver los tipos de errores más habituales en el uso de la traducción automática chino>español con el objetivo de enseñar esos errores durante el aprendizaje de español como lengua extranjera. La traducción automática está acaparando un gran interés científico e investigador en los últimos años. Con el desarrollo tecnológico y el acelerado ritmo de vida cotidiana se ha producido un aumento de demanda en el uso de traducciones automáticas por parte de los estudiantes chinos que aprenden el idioma español. Sin embargo, el resultado de los sistemas de la traducción automática a menudo no otorga sentido ni coherencia a las palabras, y genera muchos errores. Los estudiantes deberían aprender a contrarrestar de manera correcta los errores de la traducción automática. Muchos autores han identificado los principales errores producidos por la traducción automática (Hutchins y Somers, 1992; Sitman y Piñol, 1999; Mercedes, 2002; Giménez, 2009; Dorr, 1993; Arnold, 2003; Bennett, 1993; Yang y Lange, 2003; Lindop y Tsujii, 1991; Tsujii y Fujita, 1991; y Whitelock, 1992, etc.).

Mi presente trabajo consiste en el análisis crítico y contrastivo de los errores que se cometen en la traducción automática de chino al español, en el uso de los materiales de prensa china. La prensa, por sus características tales como la periodicidad, la novedad, el interés y la objetividad es un tipo de texto muy traducido con traducción automática. El propósito de esta investigación es hacer una lista de los errores comunes de la traducción automática en prensa. Siguiendo una metodología descriptiva, se corregirán las traducciones automáticas siguiendo las recomendaciones de la industria para posesición. Se marcará los errores en diferentes colores y cada color representará un tipo de error. Por medio de esa posesición, se podrá elaborar una lista preliminar de los errores más habituales en traducción automática de chino al español ya sean errores semánticos, léxicos, sintácticos, morfológicos, ortográficos y otros errores del español. Se quiere observar si los errores de traducción automática son de todo tipo o si se producen un tipo de errores más que otros.

En el futuro la autora empleará los tipos de error identificados en este trabajo para elaborar materiales didácticos para aprender español con actividades de traducción automática. Actualmente no existen investigaciones sobre el uso de la traducción automática como herramienta para aprender español por lo que esta investigación tiene un gran sentido a nivel académico.

Los errores de la traducción automática chino > español

II Congreso Internacional T3L: Tradumática, Tecnologías de la Traducción y Localización

"Las traductoras y la traducción automática"

IX Congreso Internacional sobre Traducción, Departament de Traducció i d'Interpretació i d'Estudis de l'Àsia
Oriental

10-11 de octubre de 2016 - Universitat Autònoma de Barcelona

Autora: Hong Zhang (UAB)

Introducción

Justificación

Objetivos

Metodología

Resultados y discusiones

Conclusiones

Introducción

Justificación

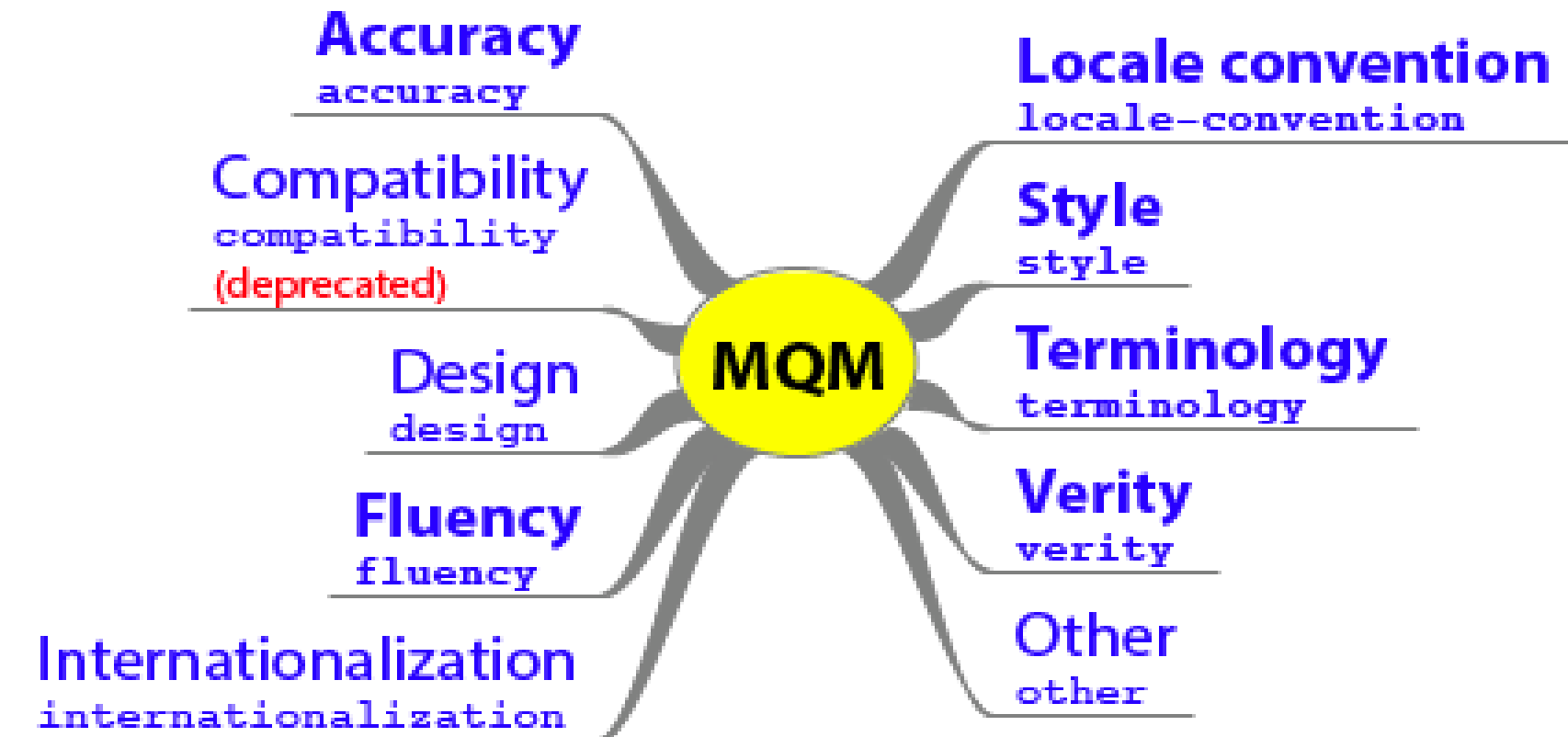
Objetivos

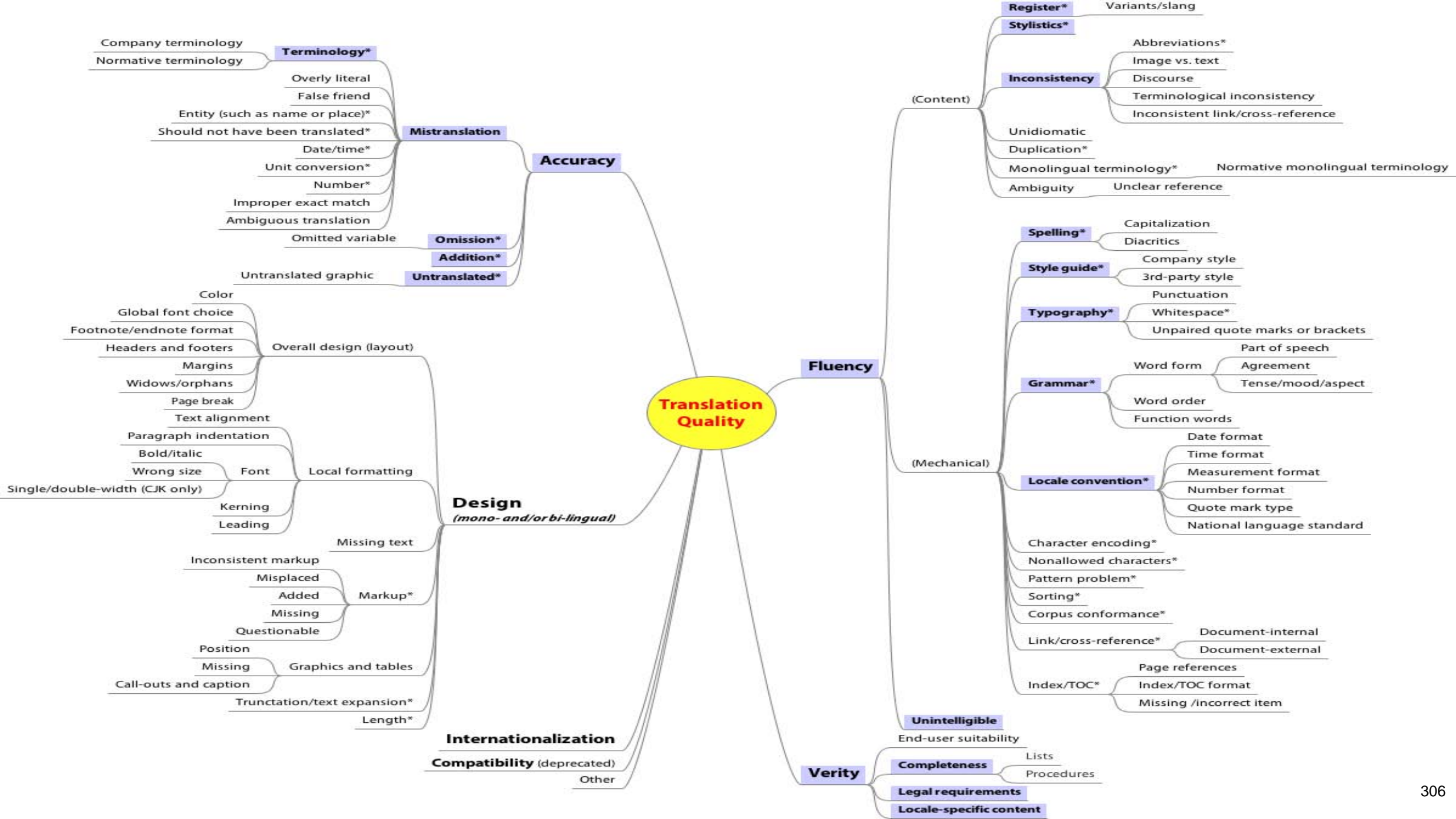
Metodología

Resultados y discusiones

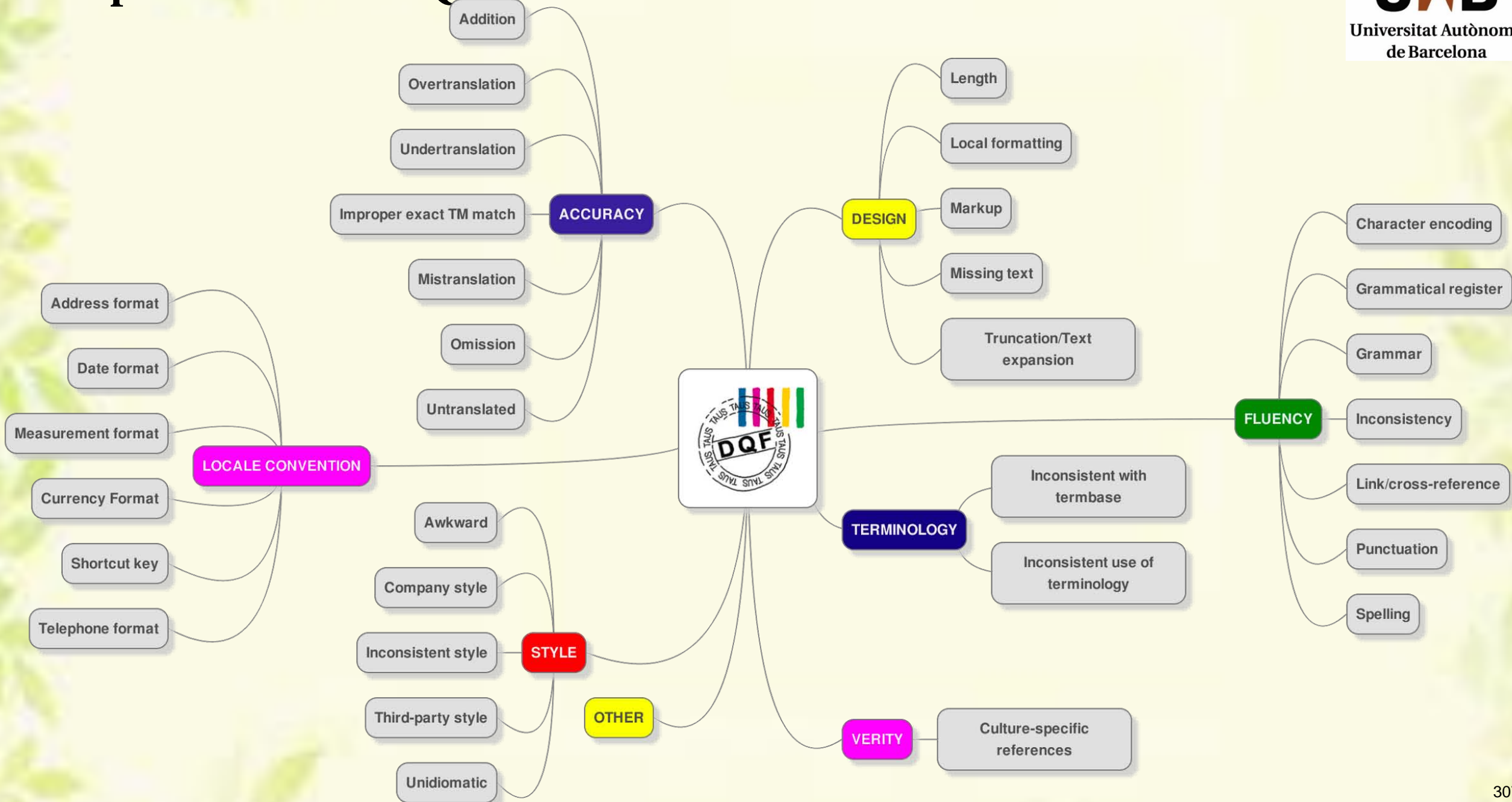
Conclusiones

Tipos de errores en: Multidimensional Quality Metrics (MQM)





Tipos de errores DQF de TAUS.net



Introducción	Justificación	Objetivos
Metodología	Resultados y discusiones	Conclusiones

La globalización y el incremento del comercio mundial genera la necesidad de producir más traducciones. Hemos analizado las diferentes traducciones automáticas obtenidas con Google Traductor, Bing y Baidu (el sistema gratuito más utilizado en China) con la finalidad de observar cuál de estos tres motores proporciona una mejor traducción. También se han analizado los diferentes tipos de errores.

Introducción

Justificación

Objetivos

Metodología

Resultados y discusiones

Conclusiones

Objetivo específico

- Evaluar los traductores automáticos más efectivos para la traducción del idioma chino > español.
- Analizar los tipos de errores de la traducción automática del idioma chino a español de la prueba realizada.
- Elaborar una lista de los errores encontrados en la traducción automática analizada, con el fin de empezar a elaborar una lista de errores que se puedan considerar específicos de la traducción automática.

Introducción	Justificación	Objetivos
Metodología	Resultados y discusiones	Conclusiones

En el desarrollo del presente trabajo se ha realizado un análisis descriptivo de la traducción automática del idioma chino a español. Para el cumplimiento de los objetivos se ha seguido los siguientes procedimientos:

1. Se han seleccionado textos escritos en el idioma chino (texto original).
2. Se han analizado tres traductores automáticos (Google, Bing y Baidu) y se ha elegido el traductor que comete menos errores.
3. La versión del traductor que cometió menos errores se comparó con una post-edición humana correcta.
4. Se identificaron los tipos de errores en la TA y se atribuyeron colores diferentes a diferentes tipos de error.
5. Se calcularon los porcentajes de los errores más habituales.

Introducción

Justificación

Objetivos

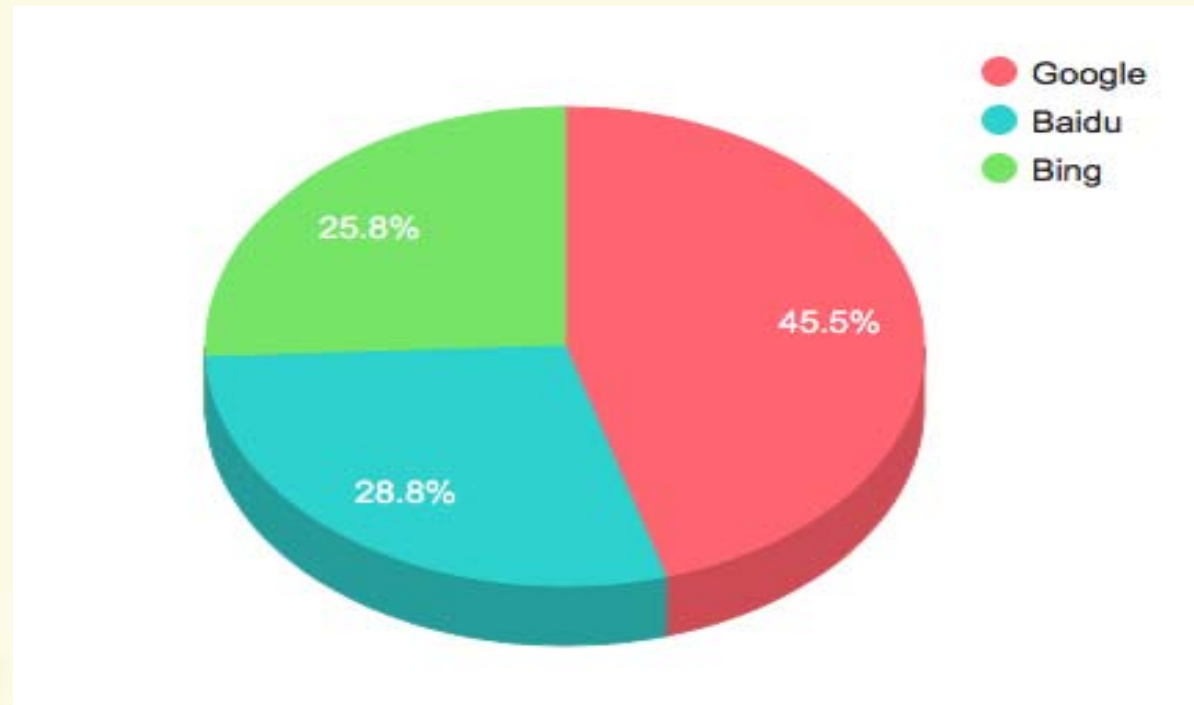
Metodología

Resultados y discusiones

Conclusiones

ANÁLISIS DE TRES TRADUCTORES AUTOMÁTICOS

Gráfico1: Comparación de motores mediante la función “Compare MT engines” de DQF Tools de TAUS



Introducción

Justificación

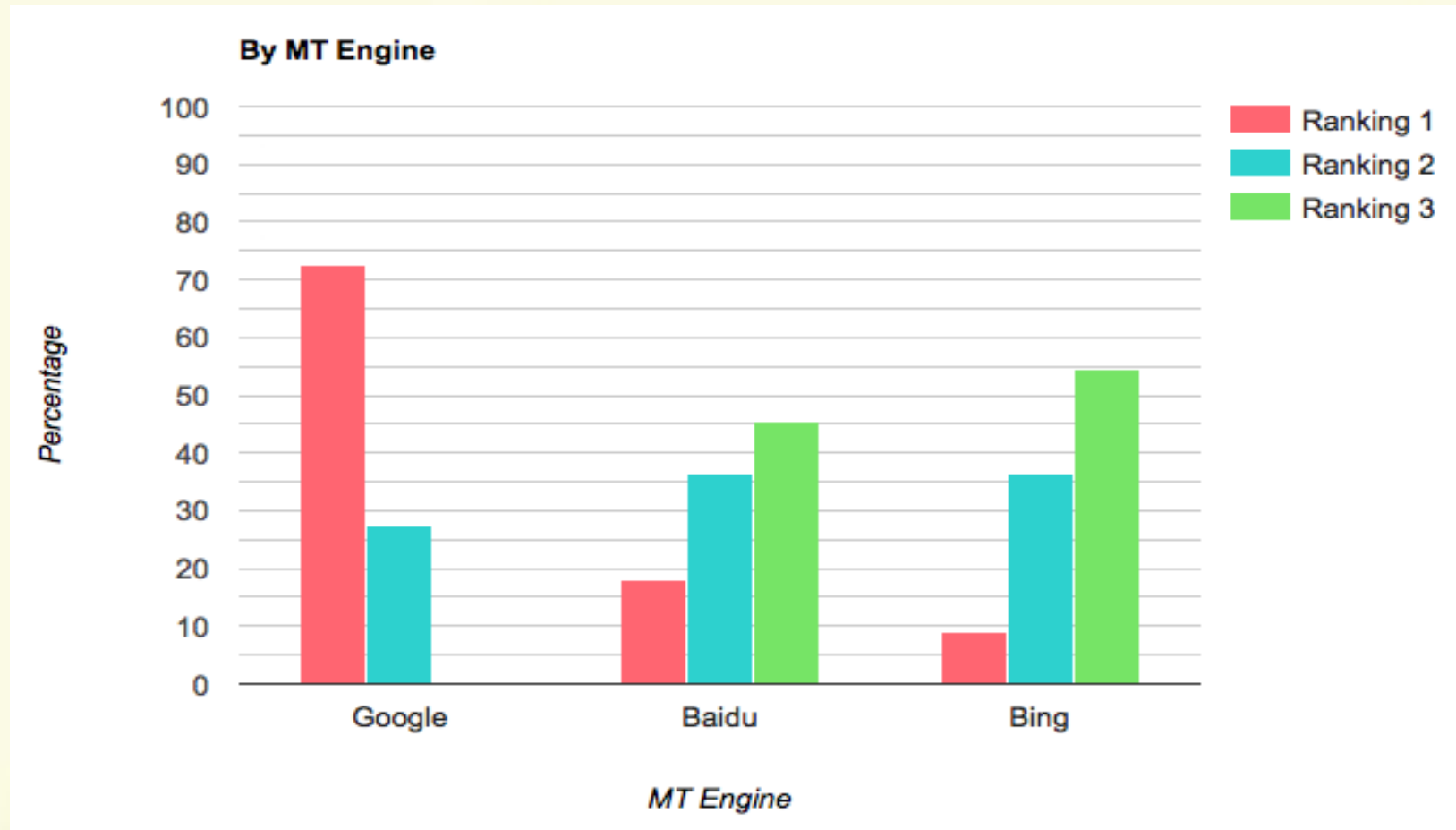
Objetivos

Metodología

Resultados y discusiones

Conclusiones

Gráfico 2: Ranking de los traductores automáticos



Introducción

Justificación

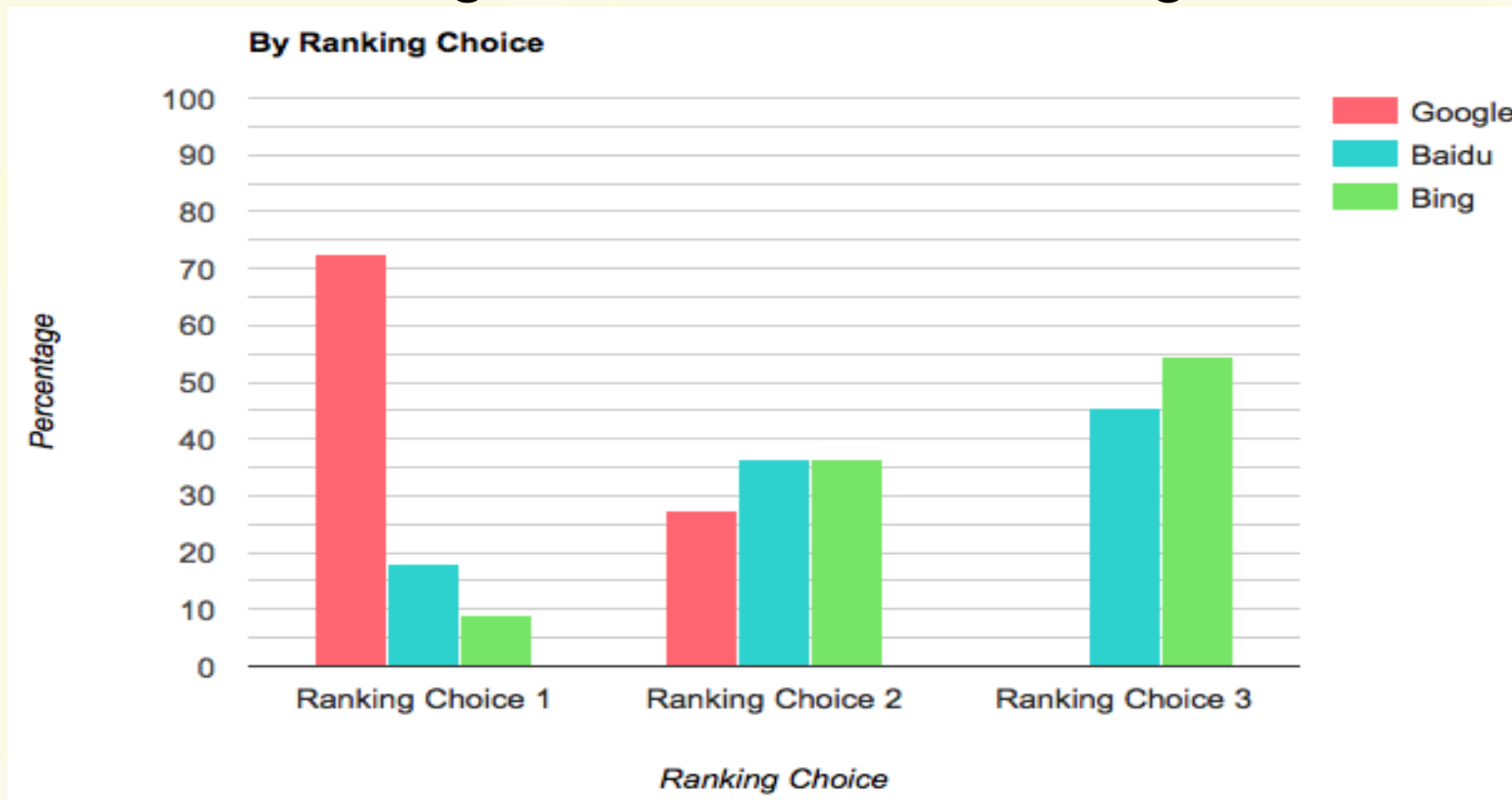
Objetivos

Metodología

Resultados y discusiones

Conclusiones

Gráfico 3: Ranking de los traductores más elegidos



Ranking 1: Motor cuyas traducciones eran las preferidas en primer lugar.

Ranking 2: ... en segundo lugar.

Ranking 3: en tercer lugar.

Introducción	Justificación	Objetivos
Metodología	Resultados y discusiones	Conclusiones

TEXTO ELEGIDO PARA ANALIZAR ERRORES DEL CHINO A ESPAÑOL

西班牙塔霍河流经西班牙腹地，经葡萄牙注入大西洋，全长910公里，水量丰沛。塞古拉河是西班牙东南部注入地中海的一条河流，流域内土壤肥沃，气候温和，农业增产潜力大，但流域内1967年缺水4.2亿立方米，2000年将缺水23.6亿立方米。因此决定从塔霍河向南调水至塞古拉河流域。每年调水10亿立方米(平均流量33立方米/秒)，除保证工业和居民用水外，还可增加灌溉面积90万公顷。

该调水工程输水道，总长286公里，须穿过两座分水岭和跨越一些河谷。该输水道包括1公里长压力钢管，总长69公里的15条隧洞，总长11公里的3条渡槽、总长160公里的渠道，中间有一段利用湖卡尔河原有水库(长45公里)作为输水道。在塔霍河引水处，需提水抬高水位260米，然后自流输水。

引水处修建博拉尔克水库(有效库容0.224亿立方米)。在该处建一座抽水蓄能电站，装4台水泵水轮机组，装机容量20万千瓦。抽水量66立方米/秒，除调水需要33立方米/秒外，其余水蓄于高山上的反调节水库中(有效库容0.48亿立方米)。

峰荷时放水发电，最大发电流量99立方米/秒。发电尾水放回博拉尔克水库

Introducción

Justificación

Objetivos

Metodología

Resultados y discusiones

Conclusiones

COMPARACION DE TEXTOS

1. Texto original

西班牙塔霍河流经西班牙腹地，经葡萄牙注入大西洋，全长910公里，水量丰沛。

2. Texto con traducción automática

España río Tajo fluye a través del corazón de España y Portugal en el Océano Atlántico después de una longitud total de 910 km, el agua abundante.

3. La propuesta de post-edición

- El río Tajo atraviesa el centro de España y Portugal, después realiza un recorrido de una longitud de 910 km, y su abundante agua desemboca en el Océano Atlántico.

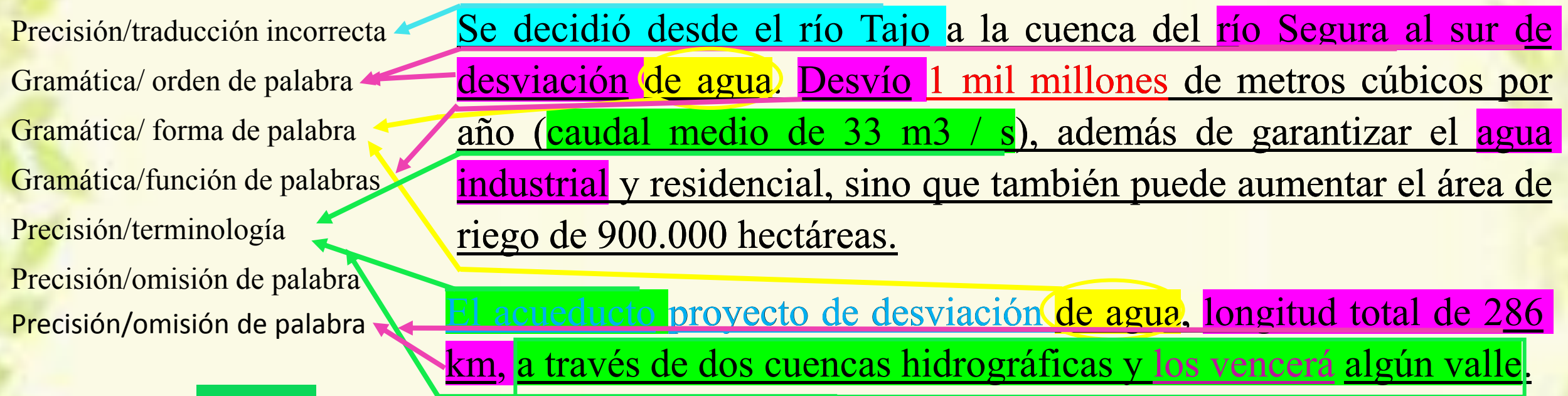
Introducción	Justificación	Objetivos
Metodología	Resultados y discusiones	Conclusiones

Identificación de los tipos de errores en la TA

Gramática/concordancia	<p>España río Tajo fluye a través del corazón de España y Portugal en el Océano Atlántico después de una longitud total de 910 km, el agua abundante.</p>
Precisión/puntuación/traducción literal	
Precisión/terminología	
Gramática/concordancia	
Tipografía/puntuación	<p>Río Segura en el sureste de España en el Mediterráneo en una cuenca hidrográfica suelo fértil, clima templado, el potencial de producción agrícola, pero la falta de agua en la cuenca en 1967 de 420 millones de metros cúbicos en 2000 a 2,36 millones de metros cúbicos de agua.</p>
Fluidez/Gramática/orden de la palabra	
Gramática/función de palabras	
Precisión/traducción incorrecta	
Gramática/forma de palabra	
Precisión/omisión (semántico)	
Precisión/número (sintáctico)	
Precisión/traducción literal	

Introducción	Justificación	Objetivos
Metodología	Resultados y discusiones	Conclusiones

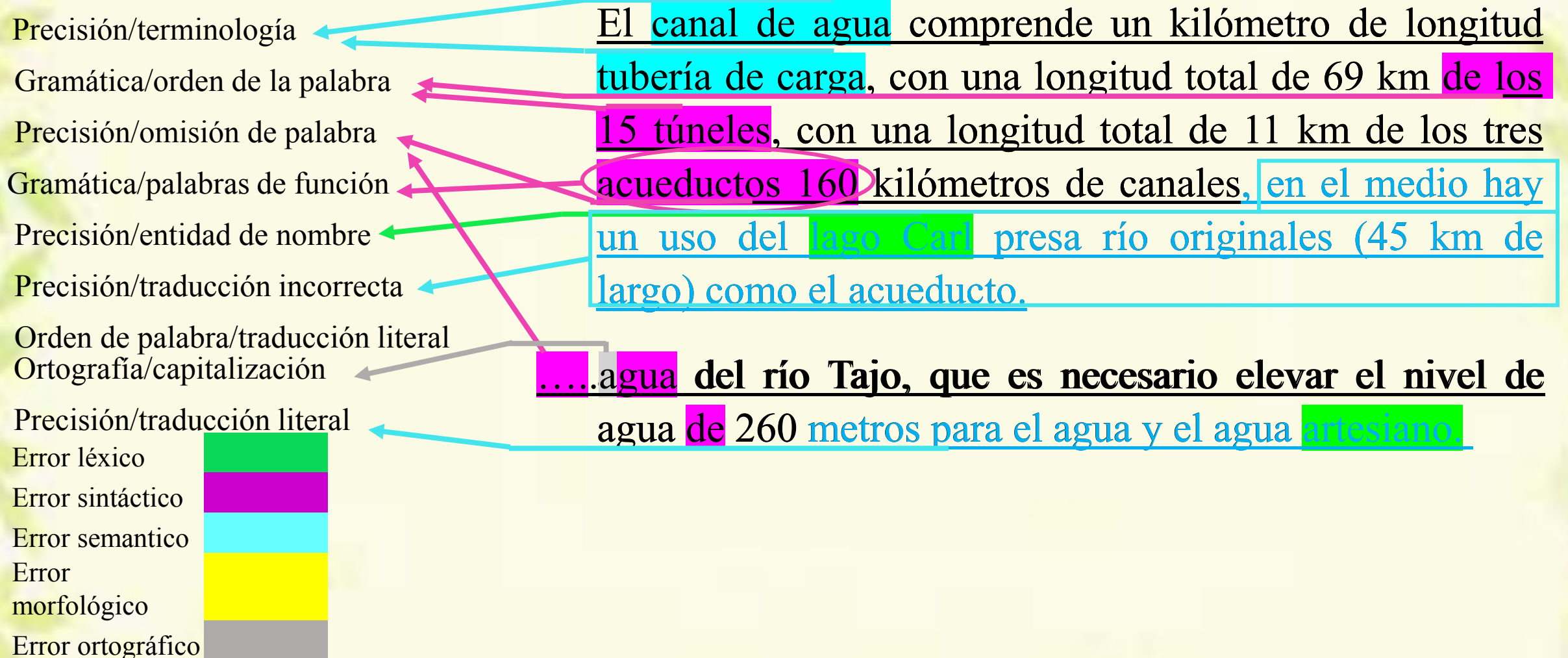
Identificación de los tipos de errores en la T.A.



Error léxico	
Error sintáctico	
Error semántico	
Error morfológico	
Error ortográfico	

Introducción	Justificación	Objetivos
Metodología	Resultados y discusiones	Conclusiones

Identificación de los tipos de errores en la T.A.



Ortografía/capitalización

Precisión/entidad de nombre

Precisión/traducción incorrecta

Precisión/numero (sintáctico)

Fluidez/ambigüedad

Precisión/terminología

Gramática/palabras de función

Gramática/forma de palabra

Precisión/omisión de palabra

Error léxico

Error sintáctico

Error semántico

Error

morfológico

Error ortográfico

agua Bora Berk embalse en la construcción (la capacidad efectiva de 022.4 millones de metros cúbicos).

Donde la construcción de una central eléctrica de acumulación por bombeo, instalado cuatro grupos turbina-bomba, la capacidad instalada de 200.000 kilovatios. La capacidad de bombeo de 66 m³ / s, además de la desviación del agua requiere de 33 metros cúbicos / segundo, el resto de la reserva de agua en las montañas en el depósito anti-regulación (la capacidad efectiva de 048 millones de metros cúbicos).

....consumo de potencia de carga máxima, la potencia máxima de flujo de 99 m³ / seg. Bora Turk agua de cola generación de vuelta al depósito.

Introducción	Justificación	Objetivos
Metodología	Resultados y discusiones	Conclusiones

ANÁLISIS DE LOS ERRORES EMITIDOS POR TRADUCTOR GOOGLE

Gráfico 4: Evaluación de fluidez con DQF Tool “Evaluate Quality” de TAUS



Introducción

Justificación

Objetivos

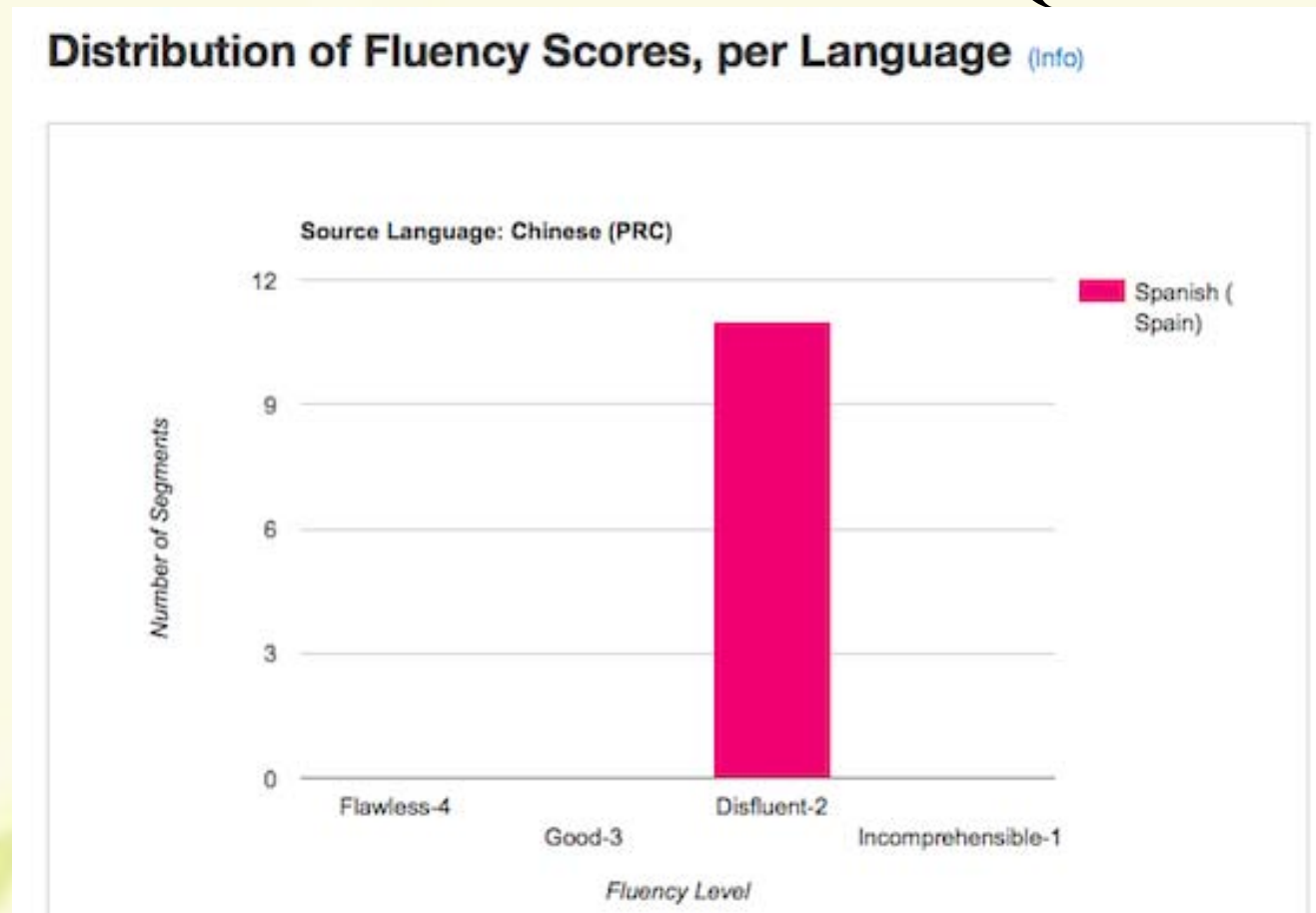
Metodología

Resultados y discusiones

Conclusiones

ANÁLISIS DE LOS ERRORES EMITIDOS POR TRADUCTOR GOOGLE

Gráfico 5: Puntuación la fluidez mediante DQF de TAUS



Introducción

Justificación

Objetivos

Metodología

Resultados y discusiones

Conclusiones

ANÁLISIS DE LOS ERRORES EMITIDOS POR TRADUCTOR GOOGLE

Gráfico 6: Evaluación de adecuación con DQF de TAUS

Average Adequacy [\(Info\)](#)

Language Pair	Average Adequacy	MT Engine	Number of Segments	Number of Words
Chinese (PRC) > Spanish (Spain)	3	Google Translate	11	476

Legend

- 4 Everything
- 3 Most
- 2 Little
- 1 None

Introducción	Justificación	Objetivos
Metodología	Resultados y discusiones	Conclusiones

ANÁLISIS DE LOS ERRORES EMITIDOS POR TRADUCTOR GOOGLE

Gráfico 7: Puntuación de la adecuación con DQF de TAUS.

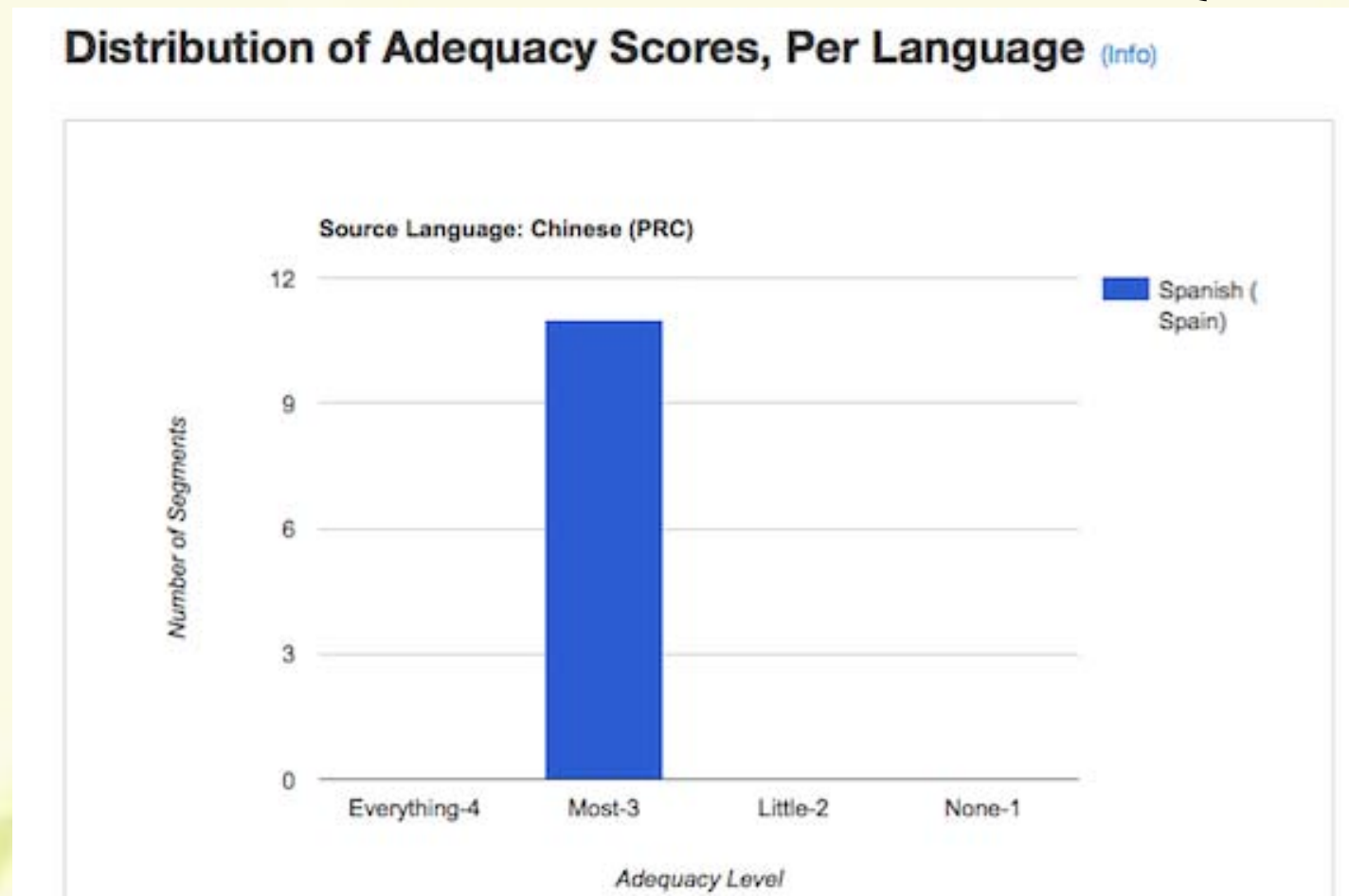


Gráfico 8: Número de los errores habituales de la traducción automática

Chinese (PRC) > Spanish (Spain)

MT Engine: Google Translate

Number of Segments: 11

Number of Words: 476

Error Type	Number of Errors
Evaluators:	Hong
Accuracy (Info)	25
Fluency errors (Info)	27
Terminology (Info)	22
Style (Info)	0
Locale convention (Info)	0
Total	74

Introducción	Justificación	Objetivos
Metodología	Resultados y discusiones	Conclusiones

Tabla 1: Lista resumen de los errores de la traducción automática analizada

Color de Identificación	Tipos de errores	Descripción
Error léxico		Precisión/terminología
		Precisión/terminología (léxico)
		Precisión/entidad de nombre
Error sintáctico		Gramática/concordancia
		Precisión/puntuación/traducción literal
		Tipografía/puntuación
		Gramática/función de palabras
		Fluidez/Gramática/ orden de palabra
		Precisión/numero (sintáctico)
		Precisión/omisión
Error semántico		Gramática/concordancia
		Fluidez/Gramática/orden de la palabra
		Precisión/traducción incorrecta (semántico que no tiene sentido)
		Precisión/omisión (semántico)
		Precisión/terminología
		Precisión/traducción incorrecta
		Fluidez/ambigüedad
Error morfológico	Error morfológico	Gramática/forma de palabra
Error ortográfico	Error ortográfico	Ortografía/capitalización

Discusiones

Sobre el motor de traducción

La herramienta Google está al alcance de todos y cuenta con beneficios altamente aprovechables (Laurenti y Dominguez, 2013: 8). En un estudio comparativo de traductores automáticos en los análisis lingüísticos de traducciones de frases y oraciones mencionan que el traductor Google en comparación con Systran y Reverso, tiene mejores resultados (Gonzales, 2010: 210). En nuestro pequeño estudio también ha mostrado los mejores resultados. No obstante, coincidimos con Amores (2016: 4) cuando afirma que:

“hay una gran variedad de herramientas que pueden automatizar en mayor o menor medida el proceso de traducción. Sin embargo, serán las necesidades últimas del usuario las que determinen qué herramienta es la más apropiada teniendo en cuenta no solamente cuestiones de calidad, sino de complejidad técnica y evaluación de los costes y beneficios asociados” (Amores, 2016: 4).

Discusiones

Sobre los tipos de errores en la traducción automática

Hemos podido constatar también lo que manifestaron Riedemann y Diégues (1998): “Se cometen una serie de errores en el plano semántico, a pesar de que de que el programa de la TA traduce a nivel de oración, es incapaz en algunos casos de extraer a partir del contexto intraoracional el verdadero sentido de los términos poliséuticos”

En la combinación chino-español, se cumple también lo afirmado por Teneche en relación con el inglés. “En el texto original del inglés traducido al español no son equivalentes en la semántica. A su vez las unidades léxicas del lenguaje no cumplen con los requerimientos semánticos” (Teneche, 2014: 104).

Introducción

Justificación

Objetivos

Metodología

Resultados y discusiones

Conclusiones

- El traductor automático que ha mostrado menos errores durante el proceso de la traducción de chino a español ha sido Google Traductor en un 45.5% superior en comparación con los otros traductores automáticos analizados, si bien el número de errores ha sido muy alto.
- Los tipos de errores más habituales en la traducción automática corresponden a la Fluidez y Precisión. A su vez en esta investigación, uno de los errores más particulares en la traducción automática de Google es el reconocimiento de las palabras en el mismo orden en la que se encuentra escrito en el idioma chino, es decir los adjetivos, sustantivos y su orden gramatical traducidos al español corresponden al texto original (idioma chino).

Introducció

Justificació

Objectivos

Metodologia

Resultados y discusiones

Conclusiones

- En la lista de errores habituales más relevantes durante la traducción, Google Traductor no ha reconocido los nombres propios y ha omitido muchas palabras. En la comparación de una frase del idioma chino traducida al español, el resultado muestra un mayor número de palabras, es decir una mayor longitud de la oración. Las puntuaciones usadas en el idioma chino, no siempre se han reflejado en el español.
- La traducción automática contenía muchos errores pero nos ha proporcionado una idea básica del significado del texto original. Ello puede ayudar al lector a entender la idea básica y no el significado total. Por esta razón será necesario la intervención humana en la mejora del texto traducido.
- Confirmamos que las herramientas DQF de TAUS son útiles para la comparación de motores de traducción y para la evaluación de traducción automática.

Recomendaciones finales

- La traducción automática satisface la demanda de traducciones de muchas personas y entidades que, en muchos casos, no pueden pagar servicios profesionales de traducción. Por este motivo recomendamos la mejora de las traducciones automáticas de chino a español.
- El uso de la Traducción Automática reduce costos y tiempo, pero debería ser obligatorio realizar una revisión de los resultados por parte de los especialistas en la traducción del idioma chino a español, y valorar dicha revisión.

¡GRACIAS POR SU ATENCION!

LA POSTEDICIÓN COMO NUEVA ESPECIALIZACIÓN TRANSDISCIPLINAR EN LA ERA DE LA INFORMACIÓN GLOBALIZADA

Coral Díez Carbajo

Universidad de Salamanca

www.usal.es

La post-edición se presenta como una solución para hacer frente a la cantidad de textos que requieren traducción hoy en día. Por lo general, hasta ahora, se ha concebido como una especialización para traductores. No obstante, como advierten investigaciones recientes (Ruano 2013, Bielsa 2005), en la era de la información, la traducción está integrada hoy en día en numerosas prácticas de redacción y procesamiento textual que permiten la circulación de información y noticias en la era global. Por esta razón, la formación en post-edición puede ser interesante para otros perfiles profesionales, como los periodistas, que, como han mostrado autores como Hernández Guerrero (2009) o Schäffner (2012), asumen operaciones de traducción en su labor diaria en prácticas como la transedición. Por ello, y basándonos en la experiencia personal, consideramos fundamental idear acciones formativas en post-edición destinadas tanto a futuros traductores como a periodistas que permitan a los estudiantes familiarizarse con los procedimientos de la post-edición para completar esta tarea de manera satisfactoria. En este sentido, en esta ponencia presentamos un curso destinado a estudiantes de traducción y periodismo con un plan de estudios organizado por semanas y concebido desde la perspectiva de las universidades de artes liberales estadounidenses. La importancia de la enseñanza de este curso va más allá del aprendizaje que puedan extraer los estudiantes que reciban formación sobre cómo post-editar correctamente y cómo obtener resultados de post-edición de mayor calidad. Así, creemos que esta propuesta permite contribuir a explorar la utilidad de la post-edición en ámbitos transdisciplinares muy diversos.

La postedición como nueva especialización transdisciplinar en la era de la información globalizada

II Congrés Internacional T3L: Tradumàtica, Tecnologies de la Traducció i Localització
"Els traductors i la traducció automàtica"

Mundo globalizado transmoderno: ubicuidad

- Necesitamos traducir para **convivir** (Ruano, 2013; Bielsa, 2015)
- Traducción para **cruzar barreras** del lenguaje y de la cultura
- Traducción **ubicua** aunque **inadvertida**
- Se desconoce que todo está **traducido, adaptado o modificado**

Mundo globalizado transmoderno: inmediatez

- Dependencia de la **inmediatez**
- Internet = mundo **interconectado** ¿barreras espacio-temporales?
- **Comunicación** del mundo transmoderno → traducción
- Necesidad de traducir con **plazos muy ajustados**



Productividad en postedición

- Hacer frente a la **cantidad** de textos
- Estudios de **productividad** (Guerberof, 2008; O'Brien, 2012)
- ¿Calidad? **Depende**: tipo de texto, profundidad con la que se postedita, combinación lingüística, experiencia del posteditor...
- Postedición: Sí, cuando tenga calidad aceptable y se tarde menos → **inmediatez**

Postedición: ¿amigo o enemigo?



- **Reticencia** de los traductores profesionales
- **¿Quitar trabajo?** Necesitamos traductores en todas las etapas de TA
- Postedición = **Nuevo nicho** de especialización (en auge gracias a la importancia de la **inmediatez**)
- Los **humanos usamos herramientas** para facilitarnos el trabajo:
 - Michael Groves (2015) *Friend or Foe? Google Translate in Language for Academic Purposes*

Postedición, ¿solo para traductores?

- Traducción está integrada en la **redacción de noticias** (Ruano, 2013; Bielsa, 2005)
- Karen Stetting (1989) → “**transedición**”: traducción + edición
- Reorganizar y redistribuir **información de otras fuentes** → “**original**”
- Ubicua e invisible
- **Postedición para transeditar noticias y contenido web**



Transedición en los medios de comunicación

- Pocos autores (Schäffner, 2012; Guerrero, 2008; Bielsa, 2016; Ruano, 2016)
 - ¿fuentes **originales**?
 - Práctica **interdisciplinar**
- **Invisible:** procesos de traducción camuflados y fragmentados



Técnicas usadas en la transedición

- **Reestructuración** del orden de los párrafos
- **Recontextualización** de datos fundamentales (fechas)
- **Adaptación** del TM al conocimiento previo del lector (**añadir o eliminar** información)
- **Objetivo:**
 - Favorecer la **comprensión**
 - **Naturalizar** el contenido (falsa percepción de “**originalidad**”)

¿Periodistas o traductores?



- Bielsa (2016) “las agencias de prensa aparecen como **vastas agencias de traducción**”
- Noticias redactadas en un idioma y **transeditadas** a los demás
- **Gran carga traductológica** pero realizada por periodistas
- **Periodistas = transeditores**
- Textos transeditados = (percibidos como) **textos originales**

¿Por qué transeditar?

- Disminuir **costes** → acuerdos entre agencias y periódicos
- Aumentar la **rapidez** de publicación → noticia “novedosa”
- Favorece la postedición en la transedición de noticias



¿Por qué posteditar las transediciones?

- Gran carga traductológica en la transedición de noticias
- Beneficiarse de la **disminución de tiempo** de trabajo
- Transedición **urgente**
- Versiones en catalán de *El Periódico* y *La Vanguardia*



El problema de la fugacidad

- TA agiliza pero el resultado **no es fugaz**
- Los textos periodísticos **sí se publican**
- Necesidad de posteditar para aportar **calidad** el TM
 - ***Light postedition*** comparando con el original → información correcta
 - **Full postedition** → partes del texto que se utilicen literalmente



Seguimiento de una noticia “original”

“Trump promete un refuerzo militar para disipar las dudas sobre su estrategia de seguridad”

El País, 8 de septiembre, 2016

Trump promete un refuerzo militar para disipar las dudas sobre su estrategia de seguridad

El republicano acusa a Clinton de tener “el gatillo fácil y ser muy inestable” en política exterior



PEPITO GRILLO

Corresponsal en
Estados Unidos



Washington - 8 SEP 2016 - 05:39 CEST

Si se tienen en cuenta las declaraciones del republicano Donald Trump y la demócrata Hillary Clinton, el [futuro presidente](#) de Estados Unidos estará incapacitado para ser el comandante en jefe de la primera potencia militar y el país se adentrará en una espiral dramática. Ambos candidatos se acusan mutuamente de carecer del juicio necesario en materia de seguridad nacional. La semana pasada lo hizo Clinton y este miércoles Trump.



📺 VÍDEOS

NEWSLETTERS



TE PUEDE INTERESAR

Fotos: El 'road trip' de 'Curiosity' por el desierto marciano



La traducción es un proceso ubicuo e invisible

El País - 8 SEPT	Fox News - 7 SEPT	Bloomberg - 7 SEPT	CBSNews - 7 SEPT	NBCNews - 8 SEPT
<p>Trump promete un refuerzo militar para disipar las dudas sobre su estrategia de seguridad</p> <p>El republicano acusa a Clinton de tener "el gatillo fácil y ser muy inestable" en política exterior</p> <p>Si se tienen en cuenta las declaraciones del republicano Donald Trump y la demócrata Hillary Clinton, el futuro presidente de Estados Unidos estará incapacitado para ser el comandante en jefe de la primera potencia militar y el país se adentrará en una espiral dramática. Ambos candidatos se acusan mutuamente de carecer del juicio necesario en materia de seguridad nacional. La semana pasada lo hizo Clinton y este miércoles Trump.</p> <p>En un discurso en Filadelfia, el magnate inmobiliario intentó contrarrestar los intentos de la ex secretaria de Estado de apelar, con un mensaje de firmeza en seguridad, a votantes conservadores. Trump prometió reforzar las Fuerzas Armadas y ser contundente con los enemigos.</p> <p>Estas fueron sus propuestas:</p> <ul style="list-style-type: none"> Pedir a los mandos militares que presenten un plan en sus primeros 30 días como presidente para derrotar y destruir al Estado Islámico (ISIS, por sus siglas en inglés) Aumentar de los 510.000 	<p>Trump unveils plan to boost US military</p> <p>Donald Trump unveiled his plan Wednesday for boosting defense spending and rebuilding the military, while calling for a new strategy to defeat the Islamic State, as part of a renewed focus by his campaign on national security.</p> <p>The Republican nominee also used the speech in Philadelphia to blast Hillary Clinton's foreign policy experience as secretary of state, which she has used as a central pillar of her campaign.</p> <p>Trump called Clinton "trigger-happy and very unstable" and pointed to the threat in Libya, Syria and Iraq.</p> <p>"This is Hillary Clinton's foreign policy legacy: failure and death," he said, before saying the recently released notes from the FBI probe into her private email server show that she is "totally unfit to be our commander-in-chief."</p> <p>For his own military policy,</p>	<p>Trump Renews Focus on Military Spending as Race Tightens</p> <p>Republican presidential nominee Donald Trump detailed his proposals for an expansion of U.S. military power as his race with Democrat Hillary Clinton tightens, in an appeal to service members and veterans who could give him a critical boost in November.</p> <p>If elected, Trump would ask Congress to lift military spending caps, increase defense spending, and seek a plan from generals to counter Islamic State in his first 30 days in office, he said in a speech Wednesday in Philadelphia.</p> <p>Trump also said he would increase the size of the army to about 540,000, the Marine Corps to 36 battalions, the navy to a number of surface ships and submarines "approaching" 350, and the Air Force to at least 1,200 fighter aircraft.</p> <p>"History shows that when America is not prepared is when the danger is by far the greatest. We want to deter, avoid, and prevent conflict," Trump said, calling Clinton "trigger-happy and very</p>	<p>Donald Trump says of Hillary Clinton "she's trigger-happy and unstable"</p> <p>PHILADELPHIA -- Donald Trump proposed more military spending in a speech here at the Union League on Wednesday, calling for an end to the mandatory cap on defense spending as a result of the 2011 Budget Control Act. This mandated deep cuts on discretionary spending on defense and government programs.</p> <p>"We want to deter, avoid and prevent conflict through our unquestioned military strength," Trump said. "We have the greatest people in the world. We have to give them the greatest equipment."</p> <p>What was left unclear by Trump's speech was how a massive increase in the defense budget would be paid for. Economists predict Trump's tax plan could deprive trillions of dollars in federal revenue over the next decade and Trump has said that he would not cut entitlements such as Medicare and Medicaid. Trump has also</p>	<p>Donald Trump Praises Putin, Hillary Clinton Defends Email</p> <p>NEW YORK — Wednesday's Commander-in-Chief Forum will be remembered as the time Donald Trump offered more praise for Russian strongman Vladimir Putin than America's own military leadership, which he described as "embarrassing."</p> <p>Hillary Clinton, while obviously more prepared and credible, found herself on defense from the outset over her handling of classified information and the military intervention in Libya she supported in the Obama administration.</p> <p>But the more explosive portion of the NBC News and MSNBC forum, however, was Trump, who in a short half-hour refused to reveal a secret plan to defeat ISIS, said the American military was "reduced to rubble," called Putin a "far more" capable leader than President Obama, defended and then walked back a tweet criticizing allowing women in the military, and revealed a new position on undocumented immigrants in the armed services.</p> <p>The forum was the first time Clinton and Trump appeared on the same stage and was as close to a dry run for the critical first presidential debate as they will get.</p> <p>Trump, who has struggled on foreign policy questions, was predictably erratic, light on details, and often inaccurate. Clinton fell short of a far higher bar set by her stature and experience, a concern in a political environment in which — rightly or not — candidates are often graded against expectations.</p>

Introducción: resumen de información

Trump promete un refuerzo militar para disipar las dudas sobre su estrategia de seguridad

El republicano acusa a Clinton de tener “el gatillo fácil y ser muy inestable” en política exterior

Si se tienen en cuenta las declaraciones del republicano Donald Trump y la demócrata Hillary Clinton, el futuro presidente de Estados Unidos estará **incapacitado para ser el comandante en jefe** de la primera potencia militar y el país se adentrará en una espiral dramática. Ambos candidatos se acusan mutuamente de carecer del juicio necesario en materia de seguridad nacional. La semana pasada lo hizo Clinton y este miércoles Trump. En un discurso en Filadelfia, el magnate inmobiliario intentó contrarrestar los intentos de la ex secretaria de Estado de apelar, con un mensaje de firmeza en seguridad, a votantes conservadores. Trump prometió reforzar las Fuerzas Armadas y ser contundente con los enemigos.

Grueso de la noticia: transedición literal

Estas fueron sus propuestas:

- Pedir a los mandos militares que presenten un plan en sus primeros 30 días como presidente para derrotar y destruir al Estado Islámico (ISIS, por sus siglas en inglés)
- Aumentar de los 510.000 actuales a 540.000 el número de soldados en activo del Ejército de tierra
- Ampliar a 36 el número de batallones de los Marines
- Aumentar a 350 los buques y submarinos de la Armada
- Ampliar a 1.200 el número de aviones de combate

Los incrementos, cuyo coste total Trump no detalló, se financiarían con la eliminación de los límites al gasto militar impuestos por el Congreso, una reforma presupuestaria y un incremento de la recaudación fiscal

En el discurso, Trump defendió aumentar el gasto militar a la vez que reiteró su reticencia intervencionista. Fue un intento de aproximación al concepto disuasorio de “paz a través de la fuerza” del ex presidente republicano Ronald Reagan, icono de la derecha estadounidense y que contribuyó al fin de la Guerra Fría.

El republicano vuelve a elogiar a Putin, del que dice que es más líder que Obama

“A diferencia de mi rival, mi política exterior enfatizará la diplomacia, no la destrucción”, dijo Trump. Acusó a la candidata demócrata de tener “el gatillo fácil y ser muy inestable”.

“A veces, parecía que no había un país en Oriente Próximo que Hillary Clinton no quisiera invadir, intervenir en él o derrocar”, dijo en referencia a su etapa como jefa de la diplomacia estadounidense en el primer mandato de Barack Obama entre 2009 y 2013. Le reprochó, por ejemplo, dejar un legado de “sufrimiento y muerte” en países como Irak, Libia y Siria.

Información complementaria: transedición libre

Trump se aleja así de la doctrina republicana, mientras que Clinton se acerca a ella al defender la semana pasada la prevalencia del llamado excepcionalismo americano.

El candidato republicano, que recibió el martes el apoyo de 88 ex altos cargos militares, busca con el discurso y su intervención, unas horas después, en un foro televisivo sobre seguridad nacional disipar las dudas que suscita en círculos conservadores su doctrina en política exterior.

La estrategia de Trump mezcla mensajes de dureza con otros aislacionistas y promesas que rompen con el consenso de las últimas décadas, como inmiscuirse menos en las injerencias rusas en Ucrania o incumplir la cláusula de defensa colectiva de la OTAN si los países miembros no aportan más fondos a la organización.

También es inusual la cercanía del republicano al presidente ruso, Vladímir Putin, al que volvió a elogiar. “Ciertamente **en ese sistema, ha sido un líder, mucho más de lo que ha sido nuestro presidente**”, dijo **Trump sobre Putin** en el coloquio organizado por la cadena NBC. El candidato, que dijo que **tendría una “muy buena relación”** con líderes extranjeros, insistió en que sería positivo para EE UU colaborar con Rusia en la lucha contra el ISIS.

Beneficios de la TA + PE en transedición

Inglés	Traducción Automática (Google Translate)	"Original" (El País)
Donald Trump says of Hillary Clinton "she's trigger-happy and unstable " (CBS News)	Donald Trump dice de Hillary Clinton, " ella es de gatillo fácil e inestable "	El republicano acusa a Clinton de tener "el gatillo fácil y ser muy inestable" [...]
<ul style="list-style-type: none"> Asking military generals to present a plan within 30 days to defeat and destroy ISIS, immediately after taking office Building an active Army of about 540,000 Building a Marine Corps based on 36 battalions Building a Navy nearing 350 surface ships and submarines Building an Air Force of at least 1,200 fighter aircraft (Fox News)	<ul style="list-style-type: none"> Pedir a los generales militares a presentar un plan dentro de los 30 días para derrotar y destruir ISIS , inmediatamente después de asumir el cargo La construcción de un ejército activo de aproximadamente 540.000 La construcción de un cuerpo de marina basado en 36 batallones La construcción de una marina de guerra a punto de 350 buques de superficie y submarinos La construcción de una fuerza aérea de aviones de combate , al menos, 1.200 	<ul style="list-style-type: none"> Pedir a los mandos militares que presenten un plan en sus primeros 30 días como presidente para derrotar y destruir al Estado Islámico (ISIS, por sus siglas en inglés) Aumentar de los 510.000 actuales a 540.000 el número de soldados en activo del Ejército de tierra Ampliar a 36 el número de batallones de los Marines Aumentar a 350 los buques y submarinos de la Armada Ampliar a 1.200 el número de aviones de combate
"There wasn't a country in the Middle East that Hillary Clinton didn't want to invade, intervene in or topple," Trump said. "She's trigger happy, and very unstable, [...]" (CBS News)	" No había un país en el Medio Oriente que Hillary Clinton no quería invadir , intervenir o derribar " dijo Trump . " Ella es el gatillo fácil , y muy inestable, [...]	Acusó a la candidata demócrata de tener "el gatillo fácil y ser muy inestable". "A veces, parecía que no había un país en Oriente Próximo que Hillary Clinton no quisiera invadir, intervenir en él o derrocar", [...]
pointed to the unrest in Libya, Syria and Iraq. "This is Hillary Clinton's foreign policy legacy -- failure and death," (Fox News)	apuntado a los disturbios en Libia , Siria e Irak . " Este es el legado de la política exterior de Hillary Clinton - el fracaso y la muerte "	Le reprochó, por ejemplo, dejar un legado de "sufrimiento y muerte" en países como Irak, Libia y Siria. [...]
Trump announced Tuesday that 88 retired U.S. generals and admirals had endorsed him, [...] participate in a forum on military issues (Bloomberg)	Trump anunció el martes que 88 generales y almirantes retirados estadounidenses le habían endosado , [...] participar en un foro sobre cuestiones militares	El candidato republicano, que recibió el martes el apoyo de 88 ex altos cargos militares, busca con el discurso y su intervención, unas horas después, en un foro televisivo sobre seguridad nacional

Transedición literal: técnicas de postedición

POSTEDICIÓN			
TIPO DE ERROR	FUENTE INGLÉS	TRADUCCIÓN AUTOMÁTICA	“ORIGINAL” <i>El País</i>
Ortotipográficos	unstable"	inestable "	inestable"
	invade,	invadir ,	invadir,
	ISIS,	ISIS ,	ISIS,
	Libya,	Libia ,	, Libia y
Verbos	Asking military generals to present	Pedir a los generales militares a presentar	Pedir a los mandos militares que presenten
	that Hillary Clinton didn't want to invade	que Hillary Clinton no quería invadir	que Hillary Clinton no quisiera invadir
Reformulaciones	Building an Air Force of at least 1,200 fighter aircraft	La construcción de una fuerza aérea de aviones de combate , al menos, 1.200	Ampliar a 1.200 el número de aviones de combate
	Building an active Army of about 540,000	La construcción de un ejército activo de aproximadamente 540.000	Aumentar de los 510.000 actuales a 540.000 el número de soldados en activo del Ejército de tierra
Elección incorrecta de significado	88 retired U.S. generals and admirals had endorsed him,	88 generales y almirantes retirados estadounidenses le habían endosado ,	que recibió el martes el apoyo de 88 ex altos cargos militares,

Transedición literal: técnicas de transedición

TRANSEDICIÓN			
TIPO DE MODIFICACIÓN	FUENTE INGLÉS	TRADUCCIÓN AUTOMÁTICA	“ORIGINAL” <i>El País</i>
Añadidos de información	destroy ISIS	destruir ISIS ,	destruir al Estado Islámico (ISIS , por sus siglas en inglés)
	Building an active Army of about 540,000	La construcción de un ejército activo de aproximadamente 540.000	Aumentar de los 510.000 actuales a 540.000 el número de soldados en activo del Ejército de tierra
Evitar repeticiones	<ul style="list-style-type: none"> • Building • Building • Building • Building 	<ul style="list-style-type: none"> • La construcción • La construcción • La construcción • La construcción 	<ul style="list-style-type: none"> • Aumentar • Ampliar • Aumentar • Ampliar
Reformulaciones	says of Hillary	dice de Hillary	acusa a Clinton
	a plan within 30 days [...] immediately after taking office	un plan dentro de los 30 días [...] inmediatamente después de asumir el cargo	un plan en sus primeros 30 días como presidente
Reordenación de información	“There wasn’t a country in the Middle East that Hillary Clinton didn’t want to invade, intervene in or topple,” Trump said. “She’s trigger happy, and very unstable, [...]”	" No había un país en el Medio Oriente que Hillary Clinton no quería invadir , intervenir o derribar " dijo Trump . " Ella es el gatillo fácil , y muy inestable , [...]	Acusó a la candidata demócrata de tener “el gatillo fácil y ser muy inestable”.
	pointed to the unrest in Libya, Syria and Iraq.	apuntado a los disturbios en Libia , Siria e Irak .	“A veces, parecía que no había un país en Oriente Próximo que Hillary Clinton no quisiera invadir , intervenir en él o derrocar ”, [...]”
	“This is Hillary Clinton’s foreign policy legacy -- failure and death ,”	" Este es el legado de la política exterior de Hillary Clinton - el fracaso y la muerte "	Le reprochó , por ejemplo, dejar un legado de “sufrimiento y muerte” en países como Irak , Libia y Siria . [...]”

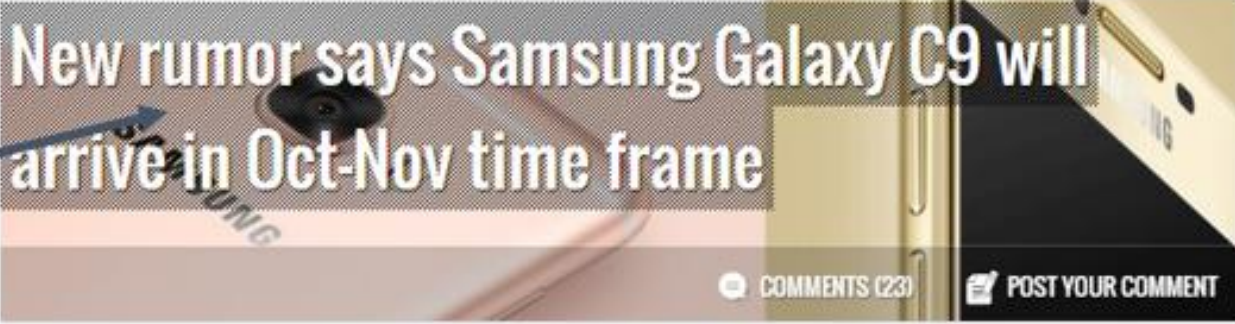


TA + PE + transedición → aumenta la velocidad

Inglés	Traducción Automática (Google Translate)	"Original" (El País)
Donald Trump says of Hillary Clinton "she's trigger-happy and unstable " (CBS News)	Donald Trump dice de Hillary Clinton, " ella es de gatillo fácil e inestable "	El republicano acusa a Clinton de tener "el gatillo fácil y ser muy inestable" [...]
<ul style="list-style-type: none"> Asking military generals to present a plan within 30 days to defeat and destroy ISIS, immediately after taking office Building an active Army of about 540,000 Building a Marine Corps based on 36 battalions Building a Navy nearing 350 surface ships and submarines Building an Air Force of at least 1,200 fighter aircraft (Fox News)	<ul style="list-style-type: none"> Pedir a los generales militares a presentar un plan dentro de los 30 días para derrotar y destruir ISIS , inmediatamente después de asumir el cargo La construcción de un ejército activo de aproximadamente 540.000 La construcción de un cuerpo de marina basado en 36 batallones La construcción de una marina de guerra a punto de 350 buques de superficie y submarinos La construcción de una fuerza aérea de aviones de combate , al menos, 1.200 	<ul style="list-style-type: none"> Pedir a los mandos militares que presenten un plan en sus primeros 30 días como presidente para derrotar y destruir al Estado Islámico (ISIS, por sus siglas en inglés) Aumentar de los 510.000 actuales a 540.000 el número de soldados en activo del Ejército de tierra Ampliar a 36 el número de batallones de los Marines Aumentar a 350 los buques y submarinos de la Armada Ampliar a 1.200 el número de aviones de combate
"There wasn't a country in the Middle East that Hillary Clinton didn't want to invade, intervene in or topple," Trump said. "She's trigger happy, and very unstable, [...]" (CBS News)	" No había un país en el Medio Oriente que Hillary Clinton no quería invadir , intervenir o derribar " dijo Trump . " Ella es el gatillo fácil , y muy inestable, [...]	Acusó a la candidata demócrata de tener "el gatillo fácil y ser muy inestable". "A veces, parecía que no había un país en Oriente Próximo que Hillary Clinton no quisiera invadir, intervenir en él o derrocar", [...]
pointed to the unrest in Libya, Syria and Iraq. "This is Hillary Clinton's foreign policy legacy -- failure and death," (Fox News)	apuntado a los disturbios en Libia , Siria e Irak . " Este es el legado de la política exterior de Hillary Clinton - el fracaso y la muerte "	Le reprochó, por ejemplo, dejar un legado de "sufrimiento y muerte" en países como Irak, Libia y Siria. [...]
Trump announced Tuesday that 88 retired U.S. generals and admirals had endorsed him, [...] participate in a forum on military issues (Bloomberg)	Trump anunció el martes que 88 generales y almirantes retirados estadounidenses le habían endosado , [...] participar en un foro sobre cuestiones militares	El candidato republicano, que recibió el martes el apoyo de 88 ex altos cargos militares, busca con el discurso y su intervención, unas horas después, en un foro televisivo sobre seguridad nacional

Beneficios de TA en la transedición libre

Inglés	Traducción Automática (Google Translate)	“Original” (El País)
<p>“Early in my term, I will also be requesting that all NATO nations promptly pay their bills. Which many are not now doing,” Trump said.</p> <p>(CBSNews)</p>	<p>"Al principio de mi mandato, también va a solicitar que todos los países de la OTAN paguen puntualmente sus facturas. La cual muchos no están haciendo ahora" dijo Trump.</p>	<p>como inmiscuirse menos en las injerencias rusas en Ucrania o incumplir la cláusula de defensa colectiva de la OTAN si los países miembros no aportan más fondos a la organización.</p>
<p>Hillary Clinton, while obviously more prepared and credible, found herself on defense from the outset over her handling of classified information and the military intervention in Libya she supported in the Obama administration [...] And they balked at her defense of the Libya intervention to oust Muammar Gaddafi</p> <p>(NBCNews)</p>	<p>Hillary Clinton, mientras que, obviamente, más preparado y creíble, se encontró en la defensiva desde el principio sobre su manejo de la información clasificada y de la intervención militar en Libia apoyó en el gobierno de Obama [...] Y ellos se opusieron a su defensa de la intervención en Libia para derrocar a Muammar Gaddafi</p>	<p>Clinton también volvió a asumir su error por votar, como senadora, a favor de la invasión de Irak de 2003. Y defendió su apoyo, como jefa de la diplomacia, a la intervención militar en Libia en 2011, que derrocó al régimen de Muammar el Gaddafi pero no estabilizó el país. La demócrata esgrimió que, de no haber intervenido, Libia se parecería hoy más a Siria.</p>
<p>Trump said of Putin. "It's a very different system and I don't happen to like the system, but certainly, in that system, he's been a leader, far more than our president has been a leader."</p> <p>(NBCNews)</p>	<p>Trump dijo Putin. "Es un sistema muy diferente y no resulta que como el sistema, pero sin duda, en ese sistema, que ha sido un líder, mucho más que nuestro presidente ha sido un líder."</p>	<p>También es inusual la cercanía del republicano al presidente ruso, Vladimir Putin, al que volvió a elogiar. "Ciertamente en ese sistema, ha sido un líder, mucho más de lo que ha sido nuestro presidente", dijo Trump sobre Putin en el coloquio organizado por la cadena NBC.</p>
<p>he said, before saying the recently released notes from the FBI probe into her private email server show that she is "totally unfit to be our commander-in-chief."</p> <p>(Fox News)</p>	<p>dijo, antes de decir las notas recién salido de la investigación del FBI en su programa de servidor de correo electrónico privada que está "totalmente inadecuada para ser nuestro comandante en jefe".</p>	<p>, el futuro presidente de Estados Unidos estará incapacitado para ser el comandante en jefe de la primera potencia militar</p>
<p>"predicted he would have a "very very good relationship" with Putin, and compared Putin favorably to Obama.</p> <p>(NBCNews)</p>	<p>"Predijo que tendría un" muy, muy buena relación" con Putin. Putin y comparado favorablemente a Obama.</p>	<p>El candidato, que dijo que tendría una "muy buena relación" con líderes extranjeros, insistió en que sería positivo para EE UU colaborar con Rusia en la lucha contra el ISIS</p>

Producción de contenidos web I

<p>Samsung 16 septiembre 2016 4:25 pm</p> <h2>El Samsung Galaxy C9 podría llegar a partir de octubre</h2>	 <p>COMMENTS (23) POST YOUR COMMENT</p> <p>Himanshu, 16 September, 2016</p> <p>Samsung Android Rumors</p>
<p>Gracias a recientes rumores sabemos que Samsung estaría trabajando en un nuevo dispositivo, que sería bautizado como Galaxy C9, y que podría integrar una amplia pantalla de 5,7 pulgadas. No tardaríamos mucho en conocerlo. Según una última información, este terminal vería la luz entre el mes de octubre o mes de noviembre, es decir dentro de muy pocas semanas. Además de contar con un panel tan generoso, el Galaxy C9 también dispondría de un procesador Qualcomm, entre 2 ó 3 GB de memoria RAM y sistema operativo Android 6.0 Marshmallow.</p> <p>Hace pocas semanas pudimos conocer mediante una filtración a través del popular sitio de importaciones Zaubu, la existencia de una nueva gama de Samsung (Galaxy C), que en un principio estaría formada por tres dispositivos: Galaxy C9, Galaxy C5 y Galaxy C7. Algunas fuentes aseguran que esta nueva serie tan solo se comercializaría en el mercado chino, aunque no hay que descartar en absoluto la posibilidad de que también se comercialice en Europa y, por tanto, en nuestro país. En cualquier caso, hoy volvemos a tener noticias del primero de ellos. Un nuevo rumor sostiene que Samsung podría empezar a vender el C9 a partir de octubre-noviembre. Eso sí, no revela nada de los territorios en los que aterrizaría.</p> <div><p>Meng Meng audio-visual</p><p>9月13日 18:45 from the iPad client</p><p>[Broke] to confirm the presence of the Samsung GALAXY C9 is expected to release time expected 5.7 / 6.0 inches in October to November @ice universe</p><p>+ Follow</p></div>	<p>There have already been <u>rumors</u> that Samsung is working on a 'C' series smartphone dubbed Galaxy C9. In fact, the device's existence has even been <u>confirmed by Zaubu</u>, where it was spotted last month. Now, a new rumor has revealed the launch time-frame for the phone.</p> <p>The rumor says that the handset will arrive sometime in the October-November time frame, which is basically within a couple of months from now.</p> <div><p>Meng Meng audio-visual</p><p>9月13日 18:45 from the iPad client</p><p>[Broke] to confirm the presence of the Samsung GALAXY C9 is expected to release time expected 5.7 / 6.0 inches in October to November @ice universe</p><p>+ Follow</p></div> <p>Sadly, nothing much is known about the device at this point in time, except that it will sport a 5.7-inch full HD display, something that was corroborated by the latest rumor as well. However, we expect more details to emerge as the launch date nears.</p>

Producción de contenidos web II

5 apps para cortar y editar vídeo en Android

Snip video Trimmer

La función principal de este editor es la de **recortar clips**. Es perfecto para ajustar el tamaño de los archivos de cara a compartirlos en las redes sociales o enviarlos **sin que pesen demasiado**. Además, incluye una función que nos permite **crear nuestros propios tonos de llamada** recortando las canciones que tengamos en nuestra biblioteca de audio.



Top 10 Best Android Video Editing Apps

#7. Snip Video Trimmer



This Android video editor is mainly used to trim videos. If you want to upload your videos on the Internet for sharing with others, you can use it to trim your video clips to get a better video and a smaller size so that you can easily upload the videos to the Internet. Besides, you can make a ringtone with this video editing app for Android.

TA en la transedición de contenido web

Inglés	Traducción Automática (Google Translate)	“Original” en español
<p>Snip Video Trimmer</p> <p>This Android video editor is mainly used to trim videos. If you want to upload your videos on the Internet for sharing with others, you can use it to trim your video clips to get a better video and a smaller size so that you can easily upload the videos to the Internet. Besides, you can make a ringtone with this video editing app for Android.</p>	<p>Snip Video Trimmer</p> <p>Este editor de vídeo Android se utiliza principalmente para recortar vídeos. Si desea cargar sus vídeos en Internet para compartir con los demás, se puede utilizar para recortar los clips de vídeo para obtener una mejor vídeo y un tamaño más pequeño para que pueda cargar fácilmente los vídeos a Internet. Además, se puede hacer un tono de llamada con esta aplicación de edición de vídeo para Android.</p>	<p>Snip video Trimmer</p> <p>La función principal de este editor es la de recortar clips. Es perfecto para ajustar el tamaño de los archivos de cara a compartirlos en las redes sociales o enviarlos sin que tarden demasiado. Además, incluye una función que nos permite crear nuestros propios tonos de llamada recortando las canciones que tengamos en nuestra biblioteca de audio.</p>
<p>KineMaster</p> <p>KineMaster is a full-featured video editing tool. It has multitracked timeline with full drag-n-drop support which enables you to easily import different types of media files and move them around with your finger. It's also equipped with robust video editing options and wide range of tools for improving video quality. Adding transitions, texts, voice-over, ect are also available.</p>	<p>KineMaster</p> <p>KineMaster es una herramienta de edición de vídeo con todas las funciones. Ha pistas múltiples línea de tiempo con soporte de arrastrar y soltar completa que le permite importar fácilmente diferentes tipos de archivos multimedia y moverlos con el dedo. También está equipado con robustas opciones de edición de vídeo y una amplia gama de herramientas para improving calidad de vídeo. La adición de transiciones, textos, voz en off, ect también están disponibles.</p>	<p>KineMaster</p> <p>KineMaster también se encuentra entre los mejores editores de vídeo gratuitos que podemos encontrar en Android ya que es realmente completo. Su interfaz es también muy intuitiva a la hora de trabajar con más de una pista de audio y vídeo y nos permite incluir pistas de audio, pegatinas, imágenes e incluso locuciones de voz. Tiene diferentes temas, algunos de pago, pero la mayoría son totalmente gratuitos.</p>

Nuevas líneas de investigación

- Creación de **cursos formativos interprofesionales** en los que se impartan **destrezas transversales** que un periodista / transeditor / traductor / posteditor necesita dominar para **transeditar con rapidez y calidad** y cumplir con las necesidades del mercado
- **Innovación docente** basada en otros cursos impartidos en grandes agencias de periodismo como la **CNN** (Gallardo Camacho, 2005)
- **Creación de motores** basándose en su corpus de noticias



Conclusiones

- **TA útil** en muchos nichos
- Traducción **ubícu**a: los periodistas son más traductores / transeditores de lo que piensan
- PE necesaria para aportar calidad por la **no fugacidad** (texto periodísticos)
 - **PE ligera**: veracidad de información
 - **PE completa**: transedición literal
- Comprensión lectora de **inglés**
- Sin recontextualización (web) → TA aumenta mucho la **velocidad**



Conclusiones II

- Aportar **visibilidad** a la labor del traductor
- Ser consciente de la **carga traductológica**
- **Limpiar la mala imagen de la TA**
- Aprender a posteditar = posteditar mejor → más calidad
- Queda **mucho por investigar**: son campos jóvenes
- Fase **inicial de investigación**



The background of the slide is a soft, misty landscape. It features rolling hills and mountains in shades of teal and light blue. A calm body of water, likely a lake, is visible in the foreground, reflecting the surrounding scenery. The overall atmosphere is serene and ethereal.

Muchas gracias

Coral Diez Carbajo

coral.diez@usal.es

TRADUCCIÓ AUTOMÀTICA I LLENGÜES MINORITZADES: EL CAS DEL SARD

Gianfranco Fronteddu i Adrià Martín-Mor

Università degli Studi di Cagliari i Universitat Autònoma de Barcelona

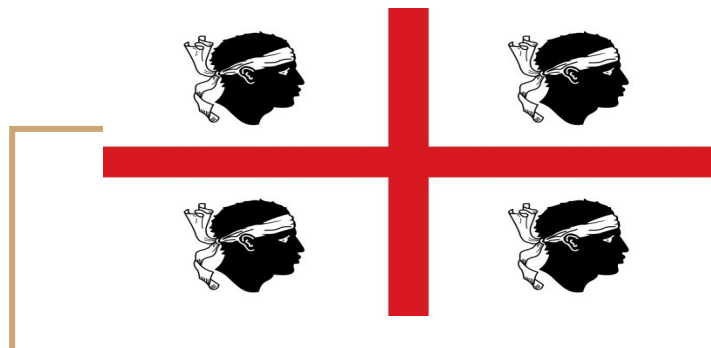
www.unica.it i www.uab.cat

La Traducció Automàtica (TA) ha esdevingut els darrers anys una eina utilitzada per professionals, investigadors i la societat en general. Malgrat que actualment els esforços en recerca i desenvolupament se centren en la millora de la qualitat, la TA en l'estat actual de la tecnologia encara pot suposar molts avenços per a tot un seguit de llengües que habitualment no tenen representació en aquests nivells. És el cas de la combinació lingüística que volem presentar, italià-sard.

El projecte que presentem neix d'una col·laboració entre universitats (la Universitat Autònoma de Barcelona i la Universitat d'Alacant) i l'empresa Prompsit, amb el finançament de Google per mitjà de la iniciativa Google Summer of Code que, cada any, permet a estudiants de grau i de postgrau el desenvolupament de projectes de codi obert durant l'estiu.

La creació d'un traductor automàtic que incorpori la llengua sarda, per les característiques d'aquesta llengua, té diverses implicacions. En primer lloc, pel fet que es tracta d'una llengua en procés d'estandardització, els recursos tant lingüístics (obres de referència) com tecnològics (corpus, correctors) són escassos. En segon lloc, aquesta escassetat, especialment de textos publicats seguint la norma estàndard, fa que calgui decantar-se per un sistema de Traducció Automàtica Basada en Regles (TABR). Apertium, inicialment un projecte públic anomenat OpenTrad, és especialment indicat per a la traducció entre parells d'idiomes de la mateixa família, i es basa en diccionaris i regles de transferència escrits amb llenguatge d'etiquetes. A més, el fet que es tracta d'un projecte lliure, disposa d'una comunitat molt nombrosa que ofereix suport als nous usuaris.

El nostre projecte demostra (i reivindica) un cop més que el perfil del traductor és vàlid per a formar part d'equips de treball per a la creació de motors de TA. En aquesta comunicació, il·lustrarem les diverses fases de creació del motor, les dificultats amb què ens hem trobat i la manera com les hem superades.



Traducció automàtica i llengües minoritzades: el cas del sard

Gianfranco Fronteddu (gfro3d@gmail.com)

Adrià Martín-Mor (adria.martin@uab.cat)

T3L, Barcelona: 11.10.2016



presentacions.tradumatica.net
/congressos



Limba sarda

Llengua sarda

- Illa de Sardenya
- Romànica, aprox. 1M parlants; 2M hab.
- UNESCO: “Definitely endangered”
- En vies d'estandardització (absència recursos)



El projecte

Su progetu

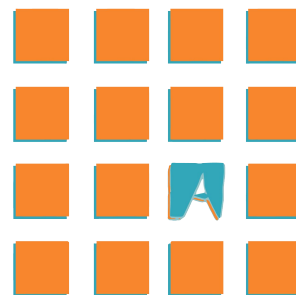
- ita > srd
- Col·laboració Tradumàtica + Prompsit + Apertium.
- Finançament de Google Summer of Code.



Google
Summer of Code
2016



Apertium





Apertium.org

- Plataforma de TABR **lliure**, originalment per a llengües romàniques. (Forcada, 2009)
- UA, 2004 (es <> ca; es <> gal).
- +40 llengües (minoritzades).

Constituïda per:

- Motor de traducció independent i modular.
- Informació lingüística per a cada combinació.
- {Eines per a la gestió de la informació lingüística}.

Per què Apertium?

Pro ite Apertium?

- Regles (no corpus).
- Lliure (participació comunitat).





Sèbera duas limbas e borta testos cun sa prataforma de tradutzione de Apertium!

italiano

español

English

Rileva sa limba in manera automàtica



sardu

català

English

Borta

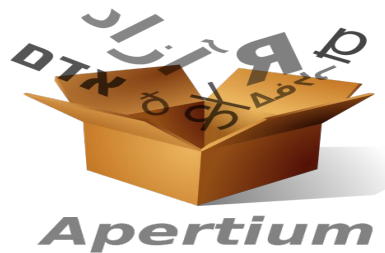


Borta unu documentu

- ☒ Sinnala sas paràulas disconnotas
- ☒ Tradutzione istantànea

La plataforma

Sa prataforma



- Codi d'etiquetes i paradigmes (independents de la llengua) › categories gramaticals
- Diccionaris morfològics en dues llengües (italià i sard) en format .xml
- Diccionari bilingüe
- Regles de transferència estructural

Apri

Salva

apertium-srd.srd.dix
~/Scrivania/Codi/apertium-srd

FileModificaVisualizzaCercaStrumentiDocumentiAiuto

<?xml version="1.0" encoding="UTF-8"?>
<!--
Dictionary:
Sections: 3
Entries: 51243
Sdefs: 65
Paradigms: 226
Last processed by: apertium-dixtools fix mono-sorted.dix apertium-srd.srd.dix.new
-->
<dictionary>

<alphabet>ÀÁÂÃÄÅÆÇÈÉÊËÌÍÎÏÑÒÓÔÕÖÙÚÛÜÝàáâãäåæçèéêëìíîïñòóôõöùúûüýÿABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyzÀÁÂÃÄÅÆÇÈÉÊËÌÍÎÏÑÒÓÔÕÖÙÚÛÜÝàáâãäåæçèéêëìíîïñòóôõöùúûüýÿÀÁÂÃÄÅÆÇÈÉÊËÌÍÎÏÑÒÓÔÕÖÙÚÛÜÝàáâãäåæçèéêëìíîïñòóôõöùúûüýÿ
alphabet>
<sdefs>
 <sdef n="n" c="Noun"/>
 <sdef n="det" c="Determiner"/>
 <sdef n="abbr" c="Abbreviation"/>
 <sdef n="acr" c="Acronym"/>
 <sdef n="percent" c=""/>
 <sdef n="adv" c="Adverb"/>
 <sdef n="preadv" c="Pre-adverb"/>
 <sdef n="predet" c="Pre-determiner"/>
 <sdef n="num" c="Numeral"/>
 <sdef n="np" c="Proper noun"/>
 <sdef n="vblex" c="Verb"/>
 <sdef n="vbser" c=""/>
 <sdef n="vbhaver" c=""/>
 <sdef n="vbmod" c="Modal verb"/>
 <sdef n="prn" c="Pronom (Pronoun)"/>
 <sdef n="pr" c="Preposition"/>
 <sdef n="adj" c="Aggettivo (Adjective)"/>
 <sdef n="cnicoo" c="Coordinating conjunction"/>

Comente funtzionat?

```

graph TD
    A[Text origen] --> B[Desformatador]
    B --> C[Analitzador morfològic]
    C --> D[Desambiguador categorial]
    D --> E[Selecció lèxica]
    E --> F[Transferència estructural]
    G[regles→] --> F
    F <--> H[Transf. lèxica]
    F --> I[Generador morfològic]
    I --> J[Post-generator]
    J --> K[Reformatador]
    K --> L[text meta]
  
```

Apertium-viewer (Italiano → Sardo)



File Download Tools Show View

Use ☐ Java ☒ C++ version

Fit

Hide intermediate

Copy

Mode: Italiano → Sardo

☒ Local ☐ Online

casa

It-proc -w /home/gianfro/Scrivania/Codi/apertium-srd-ita/ita-srd.automorf.bin

Edit source

☐ Freeze


^casa/casa<n><f><sg>\$

cg-proc -w /home/gianfro/Scrivania/Codi/apertium-srd-ita/ita-srd.rlx.bin

Edit source

☐ Freeze


apertium-tagger

☐ Freeze


^casa<n><f><sg>\$

apertium-pretransfer

☐ Freeze


It-proc -b /home/gianfro/Scrivania/Codi/apertium-srd-ita/ita-srd.autobil.bin

Edit source

☐ Freeze


^casa<n><f><sg>/domo<n><f><sg>\$

lrx-proc -m /home/gianfro/Scrivania/Codi/apertium-srd-ita/ita-srd.autolex.bin

☐ Freeze


^casa<n><f><sg>/domo<n><f><sg>\$

apertium-transfer -b /home/gianfro/Scrivania/Codi/apertium-srd-ita/apertium-srd-ita.ita-srd.t1x /home/gianfro/Scrivania/Codi/apertium-srd-ita/ita-srd.t1x.bin

Edit source

☐ Freeze


apertium-transfer: Rule 78 casa/domo [X]

^n<SN>{ ^domo<n><f><sg>\$} \$

apertium-interchunk /home/gianfro/Scrivania/Codi/apertium-srd-ita/apertium-srd-ita.ita-srd.t2x /home/gianfro/Scrivania/Codi/apertium-srd-ita/ita-srd.t2x.bin

Edit source

☐ Freeze


apertium-interchunk: Rule 1 n{ ^domo\$} [X]

^n<SN>{ ^domo<n><f><sg>\$} \$

apertium-postchunk /home/gianfro/Scrivania/Codi/apertium-srd-ita/apertium-srd-ita.ita-srd.t3x /home/gianfro/Scrivania/Codi/apertium-srd-ita/ita-srd.t3x.bin

Edit source

☐ Freeze


^domo<n><f><sg>\$

It-proc \$1 /home/gianfro/Scrivania/Codi/apertium-srd-ita/ita-srd.autogen.bin

Edit source

☐ Freeze

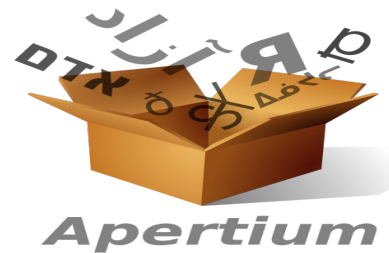

domo

It-proc -p /home/gianfro/Scrivania/Codi/apertium-srd-ita/ita-srd.autopgen.bin

☐ Freeze


domo

Monodix



- Limba Sarda Comuna
- Creació de **corpus de consulta** mitjançant:
 - blogs i revistes: “Limbas e natziones de su tempus nostru”; “Sa Gazeta”
 - Wikipedia in sardu
 - “Su Pritzipeddu”
- Estructura bàsica del diccionari sard
- Correcció diccionari ita existent

Transferència

- Bidix: 25.484 paraules



7677 <e> <p><l>discontinuidade<s n="n"/><s n="f"/></l><r>discontinuità<s n="n"/><s n="f"/></r></p><par n="ND_sp"/></e>
7678 <e> <p><l>discòrdia<s n="n"/><s n="f"/></l> <r>discordia<s n="n"/><s n="f"/></r></p></e>
7679 <e> <p><l>arresonare<s n="vblex"/></l> <r>discorrere<s n="vblex"/></r></p></e>
7680 <e> <p><l>arresonada<s n="n"/><s n="f"/></l> <r>discorso<s n="n"/><s n="m"/></r></p></e>
7681 <e> <p><l>discursu<s n="n"/><s n="m"/></l> <r>discorso<s n="n"/><s n="m"/></r></p></e>
7682 <e> <p><l>arresonu<s n="n"/><s n="m"/></l> <r>discorso<s n="n"/><s n="m"/></r></p></e>
7683 <e> <p><l>chistionu<s n="n"/><s n="m"/></l> <r>discorso<s n="n"/><s n="m"/></r></p></e>
7684 <e> <p><l>discoteca<s n="n"/><s n="f"/></l> <r>discoteca<s n="n"/><s n="f"/></r></p></e>
7685 <e> <p><l>discrèditu<s n="n"/><s n="m"/></l> <r>discredito<s n="n"/><s n="m"/></r></p></e>
7686 <e> <p><l>in/manera/discreta<s n="adv"/></l><r>discretamente<s n="adv"/></r></p></e>
7687 <e> <p><l>discretu<s n="adj"/></l> <r>discreto<s n="adj"/></r></p></e>
7688 <e r="RL"><p><l>discretu<s n="adj"/></l> <r>discreto<s n="adj"/><s n="sup"/></r></p></e>
7689 <e> <p><l>discretzionale<s n="adj"/><s n="mf"/></l><r>discrezionale<s n="adj"/><s n="mf"/></r></p></e>
7690 <e> <p><l>discrezione<s n="n"/><s n="f"/></l> <r>discrezione<s n="n"/><s n="f"/></r></p></e>
7691 <e> <p><l>discriminare<s n="vblex"/></l> <r>discriminare<s n="vblex"/></r></p></e>
7692 <e> <p><l>discriminatòriu<s n="adj"/></l> <r>discriminatorio<s n="adj"/></r></p></e>
7693 <e> <p><l>discriminatzione<s n="n"/><s n="f"/></l><r>discriminazione<s n="n"/><s n="f"/></r></p></e>
7694 <e> <p><l>dibata<s n="n"/><s n="f"/></l> <r>discussione<s n="n"/><s n="f"/></r></p></e>
7695 <e> <p><l>discussione<s n="n"/><s n="f"/></l> <r>discussione<s n="n"/><s n="f"/></r></p></e>
7696 <e> <p><l>discussu<s n="adj"/></l> <r>discusso<s n="adj"/></r></p></e>
7697 <e r="RL"><p><l>discussu<s n="adj"/></l> <r>discusso<s n="adj"/><s n="sup"/></r></p></e>
7698 <e> <p><l>discutire<s n="vblex"/></l> <r>discutere<s n="vblex"/></r></p></e>
7699 <e> <p><l>discùtere<s n="vblex"/></l> <r>discutere<s n="vblex"/></r></p></e>
7700 <e> <p><l>chistionare<s n="vblex"/></l> <r>discutere<s n="vblex"/></r></p></e>
7701 <e> <p><l>discutibile<s n="adj"/></l> <r>discutibile<s n="adj"/></r></p></e>
7702 <e r="RL"><p><l>discutibile<s n="adj"/></l> <r>discutibile<s n="adj"/><s n="sup"/></r></p><par n="mf_GD"/></e>
7703 <e> <p><l>disdeta<s n="n"/><s n="f"/></l> <r>disdetta<s n="n"/><s n="f"/></r></p></e>
7704 <e> <p><l>disdire<s n="vblex"/></l> <r>disdire<s n="vblex"/></r></p></e>
7705 <e> <p><l>disegnare<s n="vblex"/></l> <r>disegnare<s n="vblex"/></r></p></e>
7706 <e> <p><l>disignare<s n="vblex"/></l> <r>disegnare<s n="vblex"/></r></p></e>
7707 <e> <p><l>disinnare<s n="vblex"/></l> <r>disegnare<s n="vblex"/></r></p></e>
7708 <e> <p><l>disegnadore<s n="n"/></l> <r>disegnatore<s n="n"/></r></p></e>
7709 <e r="LR" c="non-preferred"><p><l>disignu<s n="n"/><s n="m"/></l><r>disegno<s n="n"/><s n="m"/></r></p></e>
7710 <e> <p><l>disinnu<s n="n"/><s n="m"/></l> <r>disegno<s n="n"/><s n="m"/></r></p></e>
7711 <e r="LR" c="non-preferred (gen. from aut. dic. comparison)"><p><l>disignu<s n="n"/><s n="m"/></l><r>disegno<s n="n"/><s n="m"/></r></p></e>
7712 <e> <p><l>disertare<s n="vblex"/></l> <r>disertare<s n="vblex"/></r></p></e>
7713 <e> <p><l>disertore<s n="n"/></l> <r>disertore<s n="n"/></r></p></e>
7714 <e> <p><l>iscontzare<s n="vblex"/></l> <r>disfare<s n="vblex"/></r></p></e>
7715 <e> <p><l>disfatista<s n="n"/></l> <r>disfattista<s n="n"/></r></p><par n="mf_GD"/></e>
7716 <e> <p><l>discontzare<s n="vblex"/></l> <r>disfatista<s n="n"/></r></p></e>

Transferència

- Bidix: 25.484 paraules
- + 35 regles de selecció lèxica




```
<!-- VITA/VIDA -->
```

```
<rule weight="1.0"> <!-- regola senza contesto per scegliere una forma -->  
    <match lemma="vita" tags="n.*"><select lemma="vida" tags="n.*"/></match>  
</rule>
```

```
<rule weight="0.1">  
    <match lemma="vita" tags="n.*"><select lemma="chintu" tags="n.*"/></match>  
</rule>
```

```
<rule weight="1.0">  
<!-- "vita alta/bassa → chintu artu/bassu -->  
    <match lemma="vita" tags="n.*"><select lemma="chintu" tags="n.*"/></match>  
    <match lemma="alto" tags="adj.*"/>  
</rule>
```

```
<!--VITE/TORONÌLLIU-->
```

```
<rule weight="1.0">  
    <match tags="num.*"/>  
    <match lemma="vite" tags="n.*"><select lemma="toronìlliù" tags="n.*"/></match>
```

Transferència

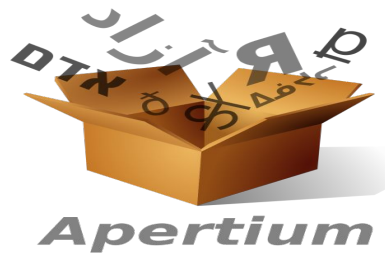
- Bidix: 25.484 paraules
- + 35 regles de selecció lèxica
- + 30 regles de desambiguació morfològica




```
171
172 SELECT N IF (0 A) (0 N) (-1 Det) (1 C A);      # dello Stato Pontificio
173 SELECT N IF (0 A) (0 N) (-1/* Pr) (1 C A);      # dello Stato Pontificio
174
175 SELECT N IF (0 A) (0 N) (-1 Det) (1 A) (1 N) (2C Pr);    # dello stato messicano di Coahuila
176 SELECT N IF (0 A) (0 N) (-1/* Pr) (1 A) (1 N) (2C Pr);  # dello stato messicano di Coahuila
177 SELECT A IF (0 A) (0 N) (-2 Det) (-1 N) (1 Pr) (NOT 1 ("da")); # dello stato messicano di Coahuila
178 SELECT A IF (0 A) (0 N) (-2/* Pr) (-1 N) (1 Pr) (NOT 1 ("da")); # dello stato messicano di Coahuila
179 REMOVE N IF (0 A) (0 N) (-2 Det) (-1 N) (1 Pr) (1 ("da")); # il titolo assunto da Filippo
180 REMOVE A IF (0 A) (0 PrfPrc) (-2 Det) (-1 N) (1 Pr) (1 ("da")); # il titolo assunto da Filippo
181
182 SELECT N IF (0 A) (0 N) (-1 Det) (1 A) (1 N) (2/* Pr); # dello stato tedesco del Baden-Württemberg
183 SELECT N IF (0 A) (0 N) (-1/* Pr) (1 A) (1 N) (2/* Pr); # dello stato tedesco del Baden-Württemberg
184 SELECT A IF (0 A) (0 N) (-2 Det) (-1 N) (1/* Pr);      # dello stato tedesco del Baden-Württemberg
185 SELECT A IF (0 A) (0 N) (-2/* Pr) (-1 N) (1/* Pr);     # dello stato tedesco del Baden-Württemberg
186 SELECT N IF (0 A) (0 N) (-1 Det) (1 A) (1 N) (2/* Det); # dello stato tedesco del Baden-Württemberg
187 SELECT N IF (0 A) (0 N) (-1/* Pr) (1 A) (1 N) (2/* Det); # dello stato tedesco del Baden-Württemberg
188 SELECT A IF (0 A) (0 N) (-2 Det) (-1 N) (1/* Det);     # dello stato tedesco del Baden-Württemberg
189 SELECT A IF (0 A) (0 N) (-2/* Pr) (-1 N) (1/* Det);    # dello stato tedesco del Baden-Württemberg
190
191 SELECT N IF (0 A) (0 N) (-1 Det) (1/* Pr);              # sullo stato dei mari
192 SELECT N IF (0 A) (0 N) (-1/* Pr) (1/* Pr);             # sullo stato dei mari
193 SELECT N IF (0 A) (0 N) (-1 Det) (1/* Det);            # sullo stato dei mari
194 SELECT N IF (0 A) (0 N) (-1/* Pr) (1/* Det);           # sullo stato dei mari
195
196 SELECT N IF (0 N) (0 PrfPrc) (-1C Det) (1 unknown);
197 SELECT N IF (0 N) (0 A) (-1C Det) (1 unknown);         # Secondo la rivista Motorrader
198 SELECT Npl IF (0 Npl) (0 PrfPrc) (-1C Num) (1 unknown);
199 SELECT Npl IF (0 Npl) (0 A) (-1C Num) (1 unknown);
200
201 #SELECT A IF (0 A) (0 N) (1 A) (-1C N);                # Regno Lombardo-Veneto (no funziona: il presidente Franco Marini)
202
203 REMOVE A IF (0 A) (0 N) (-1 Pos) (1 EOS);              # loro sorte.
204 REMOVE PrfPrc IF (0 PrfPrc) (0 N) (-1 Pos) (1 EOS);    # loro sorte.
205 REMOVE Vconj IF (0 Vconj) (0 N) (-1C A) (1 EOS);      # le più grandi collezioni.
206 REMOVE Vconj IF (0 Vconj) (0 N) (-1C A) (1 Lpar);     # due nuove canzonii (di cui una
207
208 REMOVE Sg IF (0 Vconj) (0 N) (0 Sg) (-1C Num) (NOT-1 ("uno")); # dieci canzoni
209
210 REMOVE A IF (0 A) (0 N) (-1 Pos) (1 Pr);               # il suo futuro attraverso una programmazione
211 REMOVE PrfPrc IF (0 PrfPrc) (0 N) (-1 Pos) (1 Pr);
212 REMOVE A IF (0 A) (0 N) (-1 Det) (1 Pr);              # il futuro della Ue.
213 REMOVE PrfPrc IF (0 PrfPrc) (0 N) (-1 Det) (1 Pr);
```

Regles de transferència

Regulas de trasferimentu



- **Anàlisi contrastiva**

- **Ordinals**

- (ita) terzo → (srd) su de tres

- **Imperatiu negatiu**

- (ita) non fare → (srd) non fatzas

- **Futur**

- (ita) Farò → (srd) Apo a faghene

- **Condicional**

- (ita) Farei → (srd) Dia faghene

http://wiki.apertium.org/wiki/Sardo_e_italiano/Pending_tests

Estadístiques

Istatìstìcas



- **Diccionaris:**
apertium-srd-ita.srd-ita.dix: 25.484 paraules
apertium-srd-srd.dix: 51.743 paraules
apertium-ita-ita.dix: 35.099 paraules
- **Errors:**
En un text de 2.000 paraules = 10,0%

http://wiki.apertium.org/wiki/Sardo_e_italiano/Rapporto_finale



Cobertura

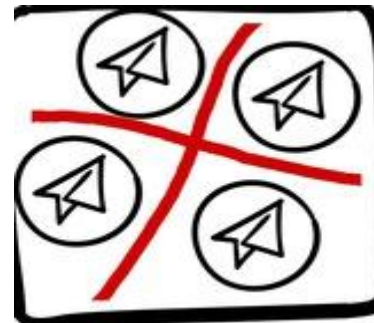
Trimmed coverage: percentatge de paraules que el traductor reconeix en un text.

Raw coverage: percentatge de paraules que l'analitzador morfològic reconeix en un text.

Resultats obtinguts amb textos de la Viquipèdia en italià per a 500, 1.000 i 2.000 paraules

Cobertura	Sardu-italianu (%)	Italianu-sardu (%)
Trimmed coverage (apertium-srd-ita)	87,8%	89,3%
Cobertura	Sardo (%)	Italianu (%)
Raw coverage (apertium-srd, apertium-ita)	88,6%	91,6%

Apertium ita-srd



- 30.08.2016 “Google Summer of Code 2016”
- 30.08.2016 Apertium ita-srd **0.9.0**
- Interfície srd www.apertium.org (Sardware)
- Col·laboració usuaris

Futur

Pro su tempus venidore

- Millora regles desambiguació morfològica i selecció lèxica.
- `srd@src?`
- `ca > srd + inverses`



Conclusions

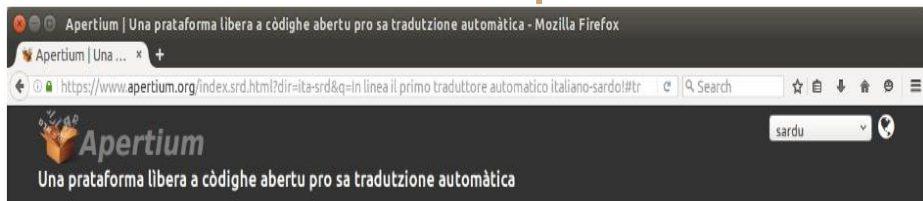
Agabbu



- Programari lliure i salvaguarda llengües minoritzades
- Apoderament comunitats + traductors
- Generació de recursos



www.apertium.org



Sèbera duas limbas e borta testos cun sa prafatorma de tradutzione de Apertium!



Gràtzias a totus
e
a medas annos!

Bibliografia

Bibliografia



- Forcada, M. (2009). “Apertium: traducció automàtica de codi obert per a les llengües romàniques”. *Linguamática 1*.
- www.blog.tradumatica.net.

WORKSHOPS

WORKSHOP: WHEN MY (TRANSLATION) MEMORY IS NOT ENOUGH: A PRACTICAL WORKSHOP ON HOW TO CREATE TRANSLATION MEMORIES FROM THE WEB AND USE THEM TOGETHER WITH MACHINE TRANSLATION.

Gema Ramírez

Promsit

www.prompsit.com

Translation memories are great: they prevent translating the same sentence twice. But, where to find them when you have no previous translations? The answer can be out there: Internet is plenty of multilingual content that can be seen as a big and powerful TM. Until now, only big players and their IT teams have been able to explore this mine, but in this workshop we will show you how you can have this power at just some clicks with a new tool: Bicrawler. We will then work on combining translation memories with machine translation to get the best of current technologies without suffering them.

Prompsit will be running this workshop. Prompsit is a translation technology company with a strong focus on empowering translators with useful tools (www.prompsit.com).

WORKSHOP: THE TAUYOU MACHINE TRANSLATION PLATFORM

Diego Bartolomé

tauyou language technology

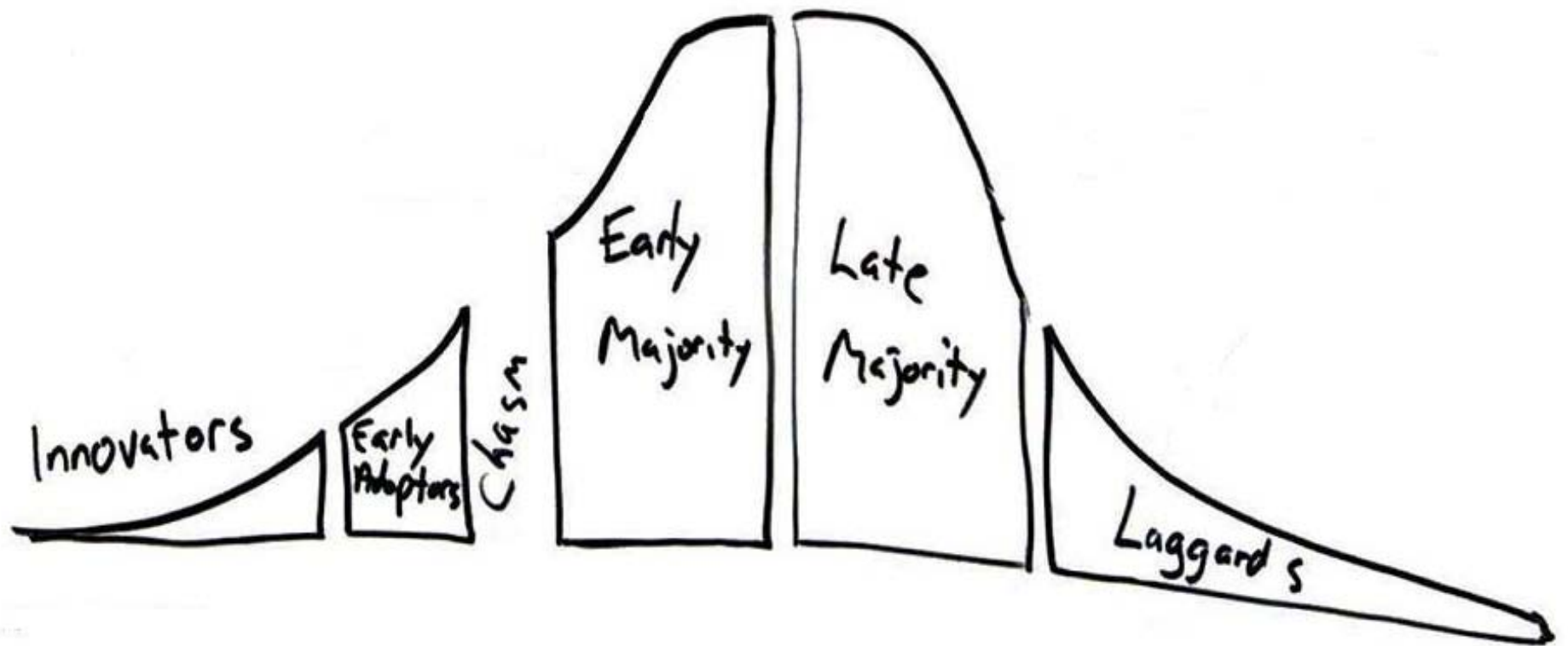
www.tauyou.com

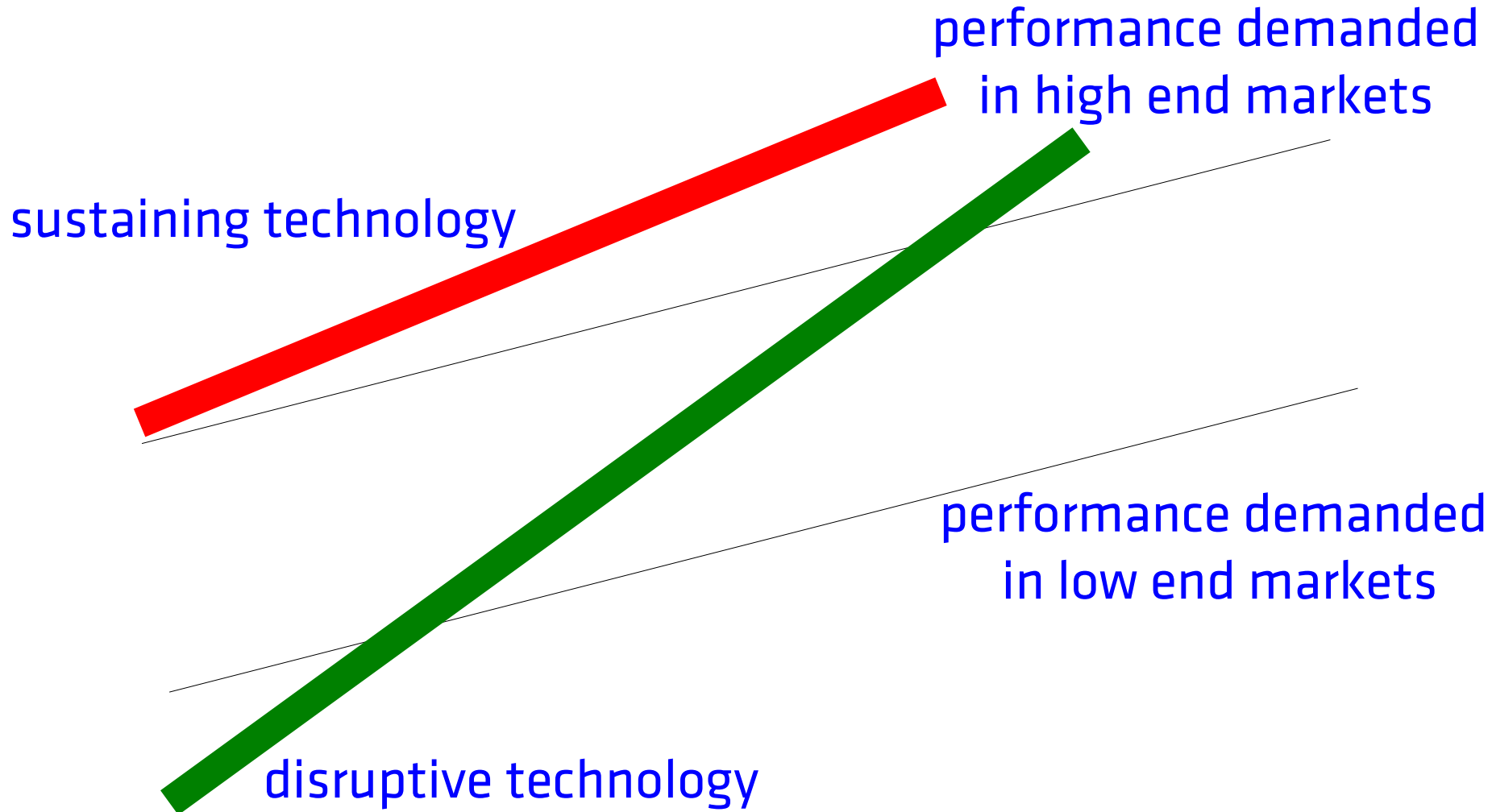
Are you tired of using Google or Microsoft Machine Translation (MT) engines where you do not have confidentiality and control? Would you like to know how to build customized MT engines to double your productivity? How can you further explore the possibilities MT offers? In this workshop, we will describe the basics of the tauyou MT platform, to then show the optimum process to achieve significant gain in productivity. Translation will be faster and better after it.

tauyou MT platform: the basics

Diego Bartolome
@diegobartolome
dbc@tauyou.com







Objectives for Machine Translation

Productivity gains

Direct cost reduction

Quality consistency

New uses for Machine Translation

Multilingual customer support

Social Media monitoring

Applications enabled by Big Data

Internet of Everything /Internet of Things

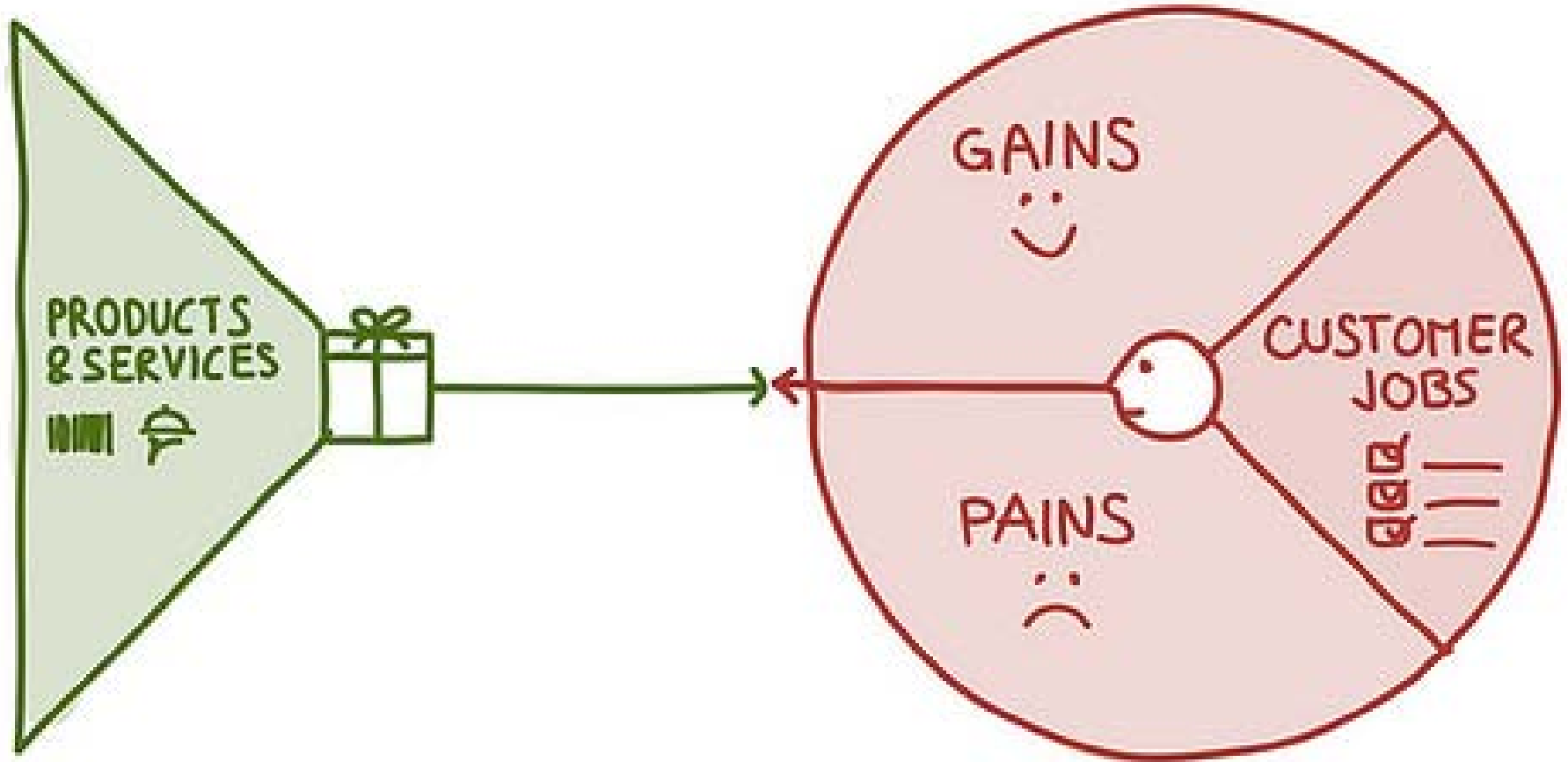
Speech-to-Speech translation

Questions

What is your experience with MT?

1. Quality Metrics
2. Cost reduction
3. Impact on Delivery Times
4. Feedback on quality
5. Your Feelings

Machine Translation Types



Google/Bing Translator vs. tauyou

Advantages

- Big(gger) data

- State-of-the-art technology

- Learning curve

Disadvantages

- Black-box

- Confidentiality

- Control

Costs of Machine Translation

Internal development – people and time

Free tools – Google + Bing

DIY solutions

Traditional pricing model

tauyou managed solution

Revenue from Machine Translation

Translation as a Service

Private Machine Translation Portal

MT of internal communication (flat rate)

....

and many others!

Questions

1. Where do you provide value now?
2. Where do you think the value will be?
3. How important is confidentiality?
4. Do you care about control?
5. How much could you invest on MT?
(time, people, money)
6. When will your solution be available?

On Language Quality

		target language																					
		EN	BG	DE	CS	DA	EL	ES	ET	FI	FR	HU	IT	LT	LV	MT	NL	PL	PT	RO	SK	SL	SV
source language	EN	–	40.5	46.8	52.6	50.0	41.0	55.2	34.8	38.6	50.1	37.2	50.4	39.6	43.4	39.8	52.3	49.2	55.0	49.0	44.7	50.7	52.0
	BG	61.3	–	38.7	39.4	39.6	34.5	46.9	25.5	26.7	42.4	22.0	43.5	29.3	29.1	25.9	44.9	35.1	45.9	36.8	34.1	34.1	39.9
	DE	53.6	26.3	–	35.4	43.1	32.8	47.1	26.7	29.5	39.4	27.6	42.7	27.6	30.3	19.8	50.2	30.2	44.1	30.7	29.4	31.4	41.2
	CS	58.4	32.0	42.6	–	43.6	34.6	48.9	30.7	30.5	41.6	27.4	44.3	34.5	35.8	26.3	46.5	39.2	45.7	36.5	43.6	41.3	42.9
	DA	57.6	28.7	44.1	35.7	–	34.3	47.5	27.8	31.6	41.3	24.2	43.8	29.7	32.9	21.1	48.5	34.3	45.4	33.9	33.0	36.2	47.2
	EL	59.5	32.4	43.1	37.7	44.5	–	54.0	26.5	29.0	48.3	23.7	49.6	29.0	32.6	23.8	48.9	34.2	52.5	37.2	33.1	36.3	43.3
	ES	60.0	31.1	42.7	37.5	44.4	39.4	–	25.4	28.5	51.3	24.0	51.7	26.8	30.5	24.6	48.8	33.9	57.3	38.1	31.7	33.9	43.7
	ET	52.0	24.6	37.3	35.2	37.8	28.2	40.4	–	37.7	33.4	30.9	37.0	35.0	36.9	20.5	41.3	32.0	37.8	28.0	30.6	32.9	37.3
	FI	49.3	23.2	36.0	32.0	37.9	27.2	39.7	34.9	–	29.5	27.2	36.6	30.5	32.5	19.4	40.6	28.8	37.5	26.5	27.3	28.2	37.6
	FR	64.0	34.5	45.1	39.5	47.4	42.8	60.9	26.7	30.0	–	25.5	56.1	28.3	31.9	25.3	51.6	35.7	61.0	43.8	33.1	35.6	45.8
	HU	48.0	24.7	34.3	30.0	33.0	25.5	34.1	29.6	29.4	30.7	–	33.5	29.6	31.9	18.1	36.1	29.8	34.2	25.7	25.6	28.2	30.5
	IT	61.0	32.1	44.3	38.9	45.8	40.6	26.9	25.0	29.7	52.7	24.2	–	29.4	32.6	24.6	50.5	35.2	56.5	39.3	32.5	34.7	44.3
	LT	51.8	27.6	33.9	37.0	36.8	26.5	21.1	34.2	32.0	34.4	28.5	36.8	–	40.1	22.2	38.1	31.6	31.6	29.3	31.8	35.3	35.3
	LV	54.0	29.1	35.0	37.8	38.5	29.7	8.0	34.2	32.4	35.6	29.3	38.9	38.4	–	23.3	41.5	34.4	39.6	31.0	33.3	37.1	38.0
	MT	72.1	32.2	37.2	37.9	38.9	33.7	48.7	26.9	25.8	42.4	22.4	43.7	30.2	33.2	–	44.0	37.1	45.9	38.9	35.8	40.0	41.6
	NL	56.9	29.3	46.9	37.0	45.4	35.3	49.7	27.5	29.8	43.4	25.3	44.5	28.6	31.7	22.0	–	32.0	47.7	33.0	30.1	34.6	43.6
	PL	60.8	31.5	40.2	44.2	42.1	34.2	46.2	29.2	29.0	40.0	24.5	43.2	33.2	35.6	27.9	44.8	–	44.1	38.2	38.2	39.8	42.1
	PT	60.7	31.4	42.9	38.4	42.8	40.2	60.7	26.4	29.2	53.2	23.8	52.8	28.0	31.5	24.8	49.3	34.5	–	39.4	32.1	34.4	43.9
	RO	60.8	33.1	38.5	37.8	40.3	35.6	50.4	24.6	26.2	46.5	25.0	44.8	28.4	29.9	28.7	43.0	35.8	48.5	–	31.5	35.1	39.4
	SK	60.8	32.6	39.4	48.1	41.0	33.3	46.2	29.8	28.4	39.4	27.4	41.8	33.8	36.7	28.5	44.4	39.0	43.3	35.3	–	42.6	41.8
	SL	61.0	33.1	37.9	43.5	42.6	34.0	47.0	31.1	28.8	38.2	25.7	42.3	34.6	37.3	30.0	45.9	38.2	44.1	35.8	38.9	–	42.7
	SV	58.5	26.9	41.0	35.6	46.6	33.3	46.6	27.4	30.9	38.9	22.7	42.0	28.2	31.0	23.7	45.6	32.2	44.2	32.7	31.3	33.5	–

Some Languages Sorted

From EN into

1) FR, ES, PT, IT

2) DE, NL, HE, DA, NO, SV

3) ZH, JA, RU

4) KR, AR, TR, HI

On Domain Quality

Who is willing to pay?

Where does your revenue come from?

What are your key skills?

What domains achieve good quality?

... Quality Order of your domains ...

Questions

1. What is your main motivation?
2. Can you try more than 1 domain?
3. Can you train at least 2 language pairs?
4. Can you pilot several MT vendors?
5. What are your expectations?

Data acquisition

OPUS corpora

<http://opus.lingfil.uu.se/>

WMT workshops

e.g. <http://www.statmt.org/wmt16/>

Multilingual websites

TAUS

Corpora building

Related vs. unrelated materials

Percentage of out-of-domain

Does mono-lingual data help?

Corpora extension with linguistic processing

Ad-hoc corpus for file translation

The more, the better?

Data cleaning

Clean translation memories

- Length, punctuation, terminology, ...

- Inconsistencies, repetitions, ...

- Segment splitting

Optimize weight of most frequent n-grams

- Validate their translations

Add out-of-domain data (optimization)

Remark

Data cleaning and selection is a key process

Just more data may harm the quality

Training strategies

One single system with all TMs

- + glossaries

- + linguistic processing input/output

- + forbidden words lists

Layered approach

Generic → domain → subdomain → client

Models optimization

Filter the translation tables

- Remove the garbage + tune weights

Optimize language models

- Adapt them to the translation purpose

Tune parameters correctly

- Tune set, test set, optimization parameters

Improve tokenization, recasing, ...

Workflow integration

Use MT as a secondary TM

Bilingual pre-translated translation files

CAT tool integration

Differentiated workflow

Continuous improvement

Qualitative

Use updated TMs in new trainings

Immediate (incremental) retraining

Rule-based automatic post-editing

Selective pre- and/or post-processing

Source content optimization

Linguistic processing notes

In the source and/or target language

Grammar checking

Entities detection

Proper nouns, alphanumeric words, ...

Compound words splitting

Sentence reordering

The Post-editor profile

Do skills needed differ from translation?

Post-editing guidelines

Full vs. light post-editing

<http://www.slideshare.net/TAUS/taus-mt-postediting-guidelines>

Compensation

Questions

Do you have the right resources to start?

Quality Metrics

SMT metrics: BLEU, NIST

Feedback from translators

Translation time vs. Post-editing time

Word Error Rate (WER) or Edit Distance

Cost reduction

Questions

Are you able to measure?



Change
before you
have to

Jack Welch

MTRADUMÀTICA

Ramon Piqué

Adrià Martín

Universitat Autònoma de Barcelona

www.tradumatica.net

La qualitat de la Traducció Automàtica (TA) es pot millorar mitjançant l'entrenament de motors estadístics a partir de corpus bilingües i monolingües.

El traductor està cridat a tenir un paper rellevant en aquest punt i entenem que cal contemplar aquests processos en la formació especialitzada del traductor.

L'MTradumàtica és una eina per a entrenar motors de TA estadístics (TAE) a partir de corpus. Es tracta d'una interfície gràfica basada en web per a Moses. L'usuari pot carregar bitextos, per a crear models de traducció, i monotextos per a crear models de llengua.

Amb l'MTradumàtica aquest procés se simplifica per mitjà d'una interfície intuïtiva i facilita l'aproximació per part dels professionals de la traducció a la gestió d'aquests processos.

MTradumàtica (English version)

The quality of machine translation (MT) can be improved by training statistical engines from monolingual and bilingual corpora.

The translator is called to play an important role in this environment and it is fundamental to include these processes in the training of specialized translators.

The MTradumàtica is a tool to train statistical MT engines (SMT) from different corpora. It is a web-based GUI for Moses. The user can load bi-texts to create translation models, and monolingual texts to create language models.

With MTradumàtica this process is simplified through an intuitive and easy interface and it facilitates translation professionals the management of these training processes.

Taller d'MTradumàtica

The 2nd International T3L Conference: Tradumatics, Translation Technologies & Localisation

9th International Conference on Translation, Department of Translation, Interpreting and East Asian Studies

Universitat Autònoma de Barcelona, 10th-11th October 2016



Grup de recerca
Tradumàtica

Adrià Martín-Mor
Ramon Piqué Huerta

MTradumàtica

MTradumàtica és una interfície gràfica d'usuari dissenyada per entrenar sistemes de traducció automàtica estadística basats en Moses.

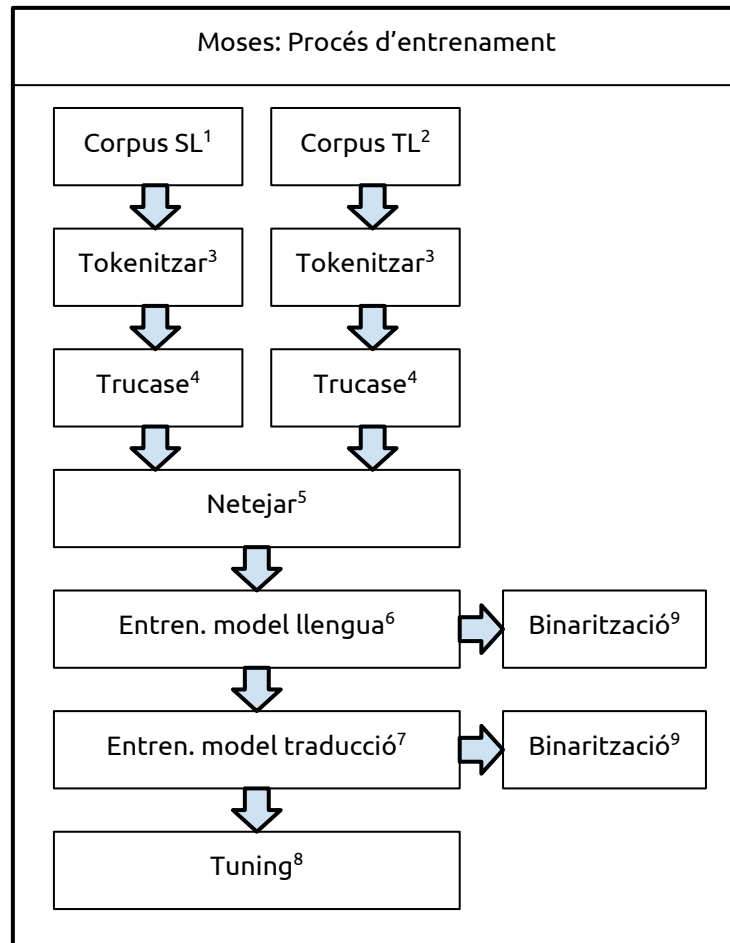
El projecte està en versió experimental.

Està concebut per a finalitats docents i de recerca.

S'ha desenvolupat en el marc del projecte ProjecTA (<http://projecTA.tradumatica.net>)



Entrenament amb Moses



MTradumàtica

Permet entrenar a partir de bitextos per crear **models de traducció** i monotextos per crear **models de llengua**.

La versió actual permet traduir en caixa de text.



MTradumàtica

El projecte preveu el següent desenvolupament:

1. Traduir arxius en **formats** diversos
2. Carregar corpus en format **TMX**
3. Crear **bitextos** a partir del web
4. Implementar un editor de **postedició**
5. Instal·lar en ordinadors personals



MTradumàtica

URL d'accés:

- <http://m.tradumatica.net>

Instal·lable en ordinadors propis (llicència lliure):

- <https://drive.google.com/file/d/0B9RM-qrF4FTIWHNLUGpTUTFLUGM/view?usp=sharing>



PRODUCTION-READY MACHINE TRANSLATION ENGINES WITH KANTANMT

Laura Casanellas Luri

KantanMT

www.kantanmt.com

Learn how to build a MT engine for a production environment in less than two hours

KantanMT is a leading provider of cloud-based Statistical Machine Translation (SMT), offering a global client base cost-effective access to the latest SMT technologies. The KantanMT.com platform is an intuitive and easy to navigate application, allowing members to build and manage SMT engines within the Amazon Web Services (AWS) cloud. The performance of SMT engines can be rapidly improved using KantanMT's unique data analysis and data visualisation technologies (KantanBuildAnalytics™), and SMT project management is simplified using KantanAnalytics™ - a SMT fuzzy-match technology used to plan, cost and schedule projects. Other features of KantanMT.com include: - PEX automatic post editing - GENTRY parsing technology - Instant Segment Re-training (KantanISR™). KantanMT seamlessly integrates with most Computer Aided Translation (CAT) tools and web applications, offering users instant access to on-demand Machine Translation. Members have reported an average increase in productivity of 60% after introducing Kantan Machine Translation into translation workflows.

New to the Kantan platform is KantanLQR™, a language quality evaluation tool that automates the process of Human Evaluation and offers a number of industry standard KPIs to choose from. Also available, is the new KantanFleet™, a set of pre-built engines in a number of verticals and language pairs.

During this workshop we aim to build a Custom Machine Translation engine through the KantanMT platform. Each attendee will be provided with access to the platform, as well as training and testing data to be used during the exercise.

We will create an engine, learn how to check the engine's initial quality using KantanBuildAnalytics, together we will identify methods to fix issues found in the training data through Pre-processing, Rejects Report and Gap Analysis features. We will learn how to keep track and compare different iterations of the engine using KantanTimeline™. Finally, we will translate a file using the recently created engine and analyse the output through Kantan's quality estimation report.

In the second part of the session, we will use KantanLQR to set up a Human Evaluation project, and finally we will review the output using the editing environment with the aim of improving the engine further to production ready status.

POST-EDITING IN A FEEDBACK FRIENDLY MT ENVIRONMENT

Blanca Vidal

Lucy Software and Services

<http://www.lucysoftware.com/espanol>

Post-Editing in a feedback friendly MT environment ...and other advantages of deep Rule-based Machine Translation Systems

The rule-based MT (RBMT) systems have often been marked as outdated and not innovative. But in reality, a deep RBMT system such as Lucy LT can yield very good results in many professional MT environments, especially for language pairs where SMT still fails to find solutions to handle language specific issues, and when not enough bi-lingual training material for the statistical training is available.

Moreover, the Lucy LT system can be quickly adapted to the customer's needs.

In the workshop will show you

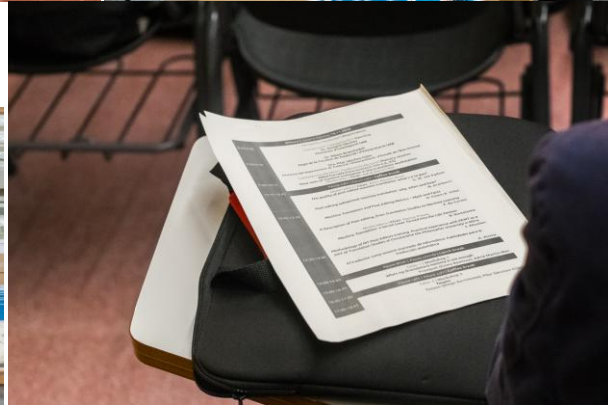
- the issues post-editors will see with MT output coming from a deep RBMT system,
- how easily the output quality could be improved with feedback from the post-editor.

Due to the interactive character of the workshop, we would recommend to limit the number of participant to max 15 persons.

PICTURES



















VIDEOS

<http://videos.tradumatica.net>